CUG-H02972-14-P004707

ANALYSIS

Edited by Christopher Kirwan

⑭

Volume 45

1985

ISSN 0003-2638

BASIL BLACKWELL · OXFORD

INDEX

		PAC
ALEXANDER, IVOR	'If' and quantification	18
BISHOP, JOHN	Causal deviancy and multiple intentions: a reply to James Montmarquet	16
BRINGSJORD, SELMER	Are there set theoretic possible worlds?	6
BRUECKNER, ANTHONY L.	Losing track of the sceptic	10
BUCKNER, D. K.	Reply to van Brakel	21
BURKE, MICHAEL B.	Spatial analogues of 'annihilation and re-creation'	2
CAIN, JAMES	Some radical consequences of Geach's logical theories	8
CARTER, W. R.	Impeccability revisited	5
CHANDLER, HUGH S.	Indeterminate people	14
CLARK, A. J.	Why Kant couldn't be an anti-realist	6
COHEN, L. JONATHAN	A problem about ambiguity in truth-theoretical semantics	12
COPELAND, B. J.	Substitutional quantification and existence	
DALE, A. J.	Is the future unreasonable?	17
DAU, PAOLO	The complex matter of denoting	19
DUDMAN, V. H.	Thinking about the future	18
EHRING, DOUGLAS	Simultaneous causation and causal chains	9
GARRETT, B. J.	A note on substance concepts	17
GARRETT, B. J.	Noonan, 'best candidate' theories and the ship of Theseus	21
GIBBINS, P. F.	Are mental events in space-time?	14
GOLDSTEIN, LAURENCE	The paradox of the liar — a case of mistaken identity	
GOLDSTEIN, LAURENCE	The title of this paper is 'quotation'	13
GORDON, DAVID	Green and dictators revisited	21
GREEN, LESLIE	What is a dictator?	12
HARDIN, C. L.	A transparent case for subjectivism?	11
HARRISON, JONATHAN	Distribution, superposition and quantum logic	20
HAZEN, ALLEN	McGinn's reply to Wright's reply to Benacerraf	5
HAZEN, ALLEN	Nominalism and abstract entities	6
HOLLIS, MARTIN	Report on ANALYSIS problem no. 20	17

Hugly, Philip *and* Sayward, Charles	Is moral relativism consistent?	40
Kennedy, Ralph	Lemmon on logical relations	89
Kirk, Robert	Davidson and indeterminacy of translation	20
Knorpp, William	Thoughtcrime	178
Kvanvig, Jonathan L.	Swain on the basing relation	153
Lockwood, Michael	Einstein, Gibbins and the unity of time	148
Lowe, E. J.	'If A and B, then A'	93
Lowe, E. J.	Reply to Wright on conditionals and transitivity	200
Madell, Geoffrey	Derek Parfit and Greta Garbo	105
Millar, Alan	Veridicality: more on Searle	120
Moore, A. W.	Set theory, Skolem's paradox and the *Tractatus*	13
Moore, F. C. T.	The martyr's dilemma	29
Morris, Thomas V.	Agnosticism	219
Noonan, H. W.	Wiggins, artefact identity and 'best candidate' theories	4
Noonan, H. W.	The only x and y principle	79
Noonan, H. W.	A note on temporal parts	151
Olding, A.	Short on teleology	158
Primorac, Igor	Hare on moral conflicts	171
Putnam, Hilary	A quick Read is a wrong Wright	203
Radford, Colin	The umpire's dilemma	109
Rasmussen, Stig Alstrup	Sainsbury on denying a Fregean conclusion	77
Read, Stephen *and* Wright, Crispin	Hairier than Putnam thought	56
Sayward, Charles	(*see* Hugly, Philip)	
Siegel, Harvey	Tarski a relativist?	75
Simpson, Paul	Lyons and tigers	169
Sinnott-Armstrong, Walter	'Ought to have' and 'could have'	44
Sorensen, Roy A.	An argument for the vagueness of 'vague'	134
Swank, Casey	Reasons, dilemmas and the logic of 'ought'	111
Tennant, Neil	Weir and those 'disproofs' I saw before me	208
Van Brakel, J.	Buckner quoting Goldstein and Davidson on quotation	73

VAN INWAGEN, PETER	On two arguments for compatibilism	161
VISION, GERALD	'I am here now'	198
WALLER, BRUCE N.	Deliberating about the inevitable	48
WASSERMAN, WAYNE	What is a fundamental ethical disagreement?	34
WEIR, ALAN	Rejoinder to Tennant	68
WRIGHT, CRISPIN	(*see* READ, STEPHEN)	

l. 45 No. 1 (New Series No. 205) January 198

105
A92-13

ANALYSIS

Edited by
Christopher Kirwan

CONTENTS

Substitutional quantification and existence	B. J. Copeland
Wiggins, artefact identity and 'best candidate' theories	H. W. Noonan
The paradox of the liar – a case of mistaken identity	Laurence Goldstein
t theory, Skolem's paradox and the *Tractatus*	A. W. Moore
Davidson and indeterminacy of translation	Robert Kirk
Spatial analogues of 'annihilation and re-creation'	Michael B. Burke
The martyr's dilemma	F. C. T. Moore
What is a fundamental ethical disagreement?	Wayne Wasserman
s moral relativism consistent?	Philip Hugly *and* Charles Sayward
Ought to have' and 'could have'	Walter Sinnott-Armstrong
Deliberating about the inevitable	Bruce N. Waller
mpeccability revisited	W. R. Carter
airier than Putnam thought	Stephen Read *and* Crispin Wright
McGinn's reply to Wright's reply to Benacerraf	Allen Hazen
hy Kant couldn't be an anti-realist	A. J. Clark
e there set theoretic possible worlds?	Selmer Bringsjord

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003–2638

Philosophical Investigations

Edited by **D.Z. Phillips**

105
Apr 12

Volume 8 No. 1 January 1985

Contents

Articles:
Was Wittgenstein A Sceptic? OSWALD HANFLING
Empiricism and the Theory of Meaning DAVID COCKBURN
The Constitutive Force of Language JAMES F. HARRIS

Critical Notice:
Colin McGinn, *The Subjective View: Secondary Qualities and Indexical Thoughts* PETER SMITH

Review:
Gordon Graham, *Historical Explanation Reconsidered* *R.F. ATKINSON*

Books Received

Published quarterly: January, April, July and October
Subscriptions Volume 8 (1985)
Institutions: *£28.00 (UK); £35.00 (Overseas); $65.55 (US); $80.30 (Canada)*
Individuals: *£12.95 (UK); £16.20 (Overseas); $30.35 (US); $37.20 (Canada)*
Single issues are available; prices on request.

Orders, with remittance, or requests for specimen copies and further details should be sent to:
Iris Taylor, Journals Department, Basil Blackwell,
108 Cowley Road, Oxford OX4 1JF, England

Basil Blackwell · Oxford · England

Collected Papers on Mathematics, Logic and Philosophy

GOTTLOB FREGE
Edited by Brian McGuinness

This book presents all Frege's mathematical writings, his reviews, fundamental articles on philosophy of language which prepared the way for his systematic works, his answers to criticism, his controversy over geometry, and all his final publications developing further his philosophy of language. Many are translated here for the first time.
416 pages, **£28.50** (0 631 12728 3)

Frege: Tradition and Influence

Edited by CRISPIN WRIGHT

The contributors to this book, a major assessment of Frege's thought and his legacy, include D. Bell, G. Currie, P. Carruthers, C. Diamond, M. Dummett, G. Gabriel, R. Hale, M. Luntley, G. McCulloch, J. McDowell, H. Noonan, J. Skorupski, J. Weiner and D. Wiggins.
270 pages, paperback **£6.95** (0 631 13831 5)

Probabilistic Metaphysics

PATRICK SUPPES

In this clear and authoritative book, Patrick Suppes argues against the classical assumptions of determinism. Denying the possibility of complete descriptions of events and laws he proposes instead a new theory of 'probabilistic' metaphysics.
256 pages, **£19.50** (0 631 13332 1)

A Radical Philosophy

AGNES HELLER

Agnes Heller's aim in this book is to offer coherent and unified answers to some fundamental philosophical and political questions: how one should think, how one should act, and how one should live. In doing so she formulates the essence of a philosophy for the radical left.
208 pages, **£17.50** (0 631 12567 1)

Basil Blackwell

Property and Political Theory

ALAN RYAN

Until a few years ago, arguments about the rights of property were central to political theory. Alan Ryan brings to life the cut and thrust of these arguments from Locke and Rousseau, through Kant and Hegel, to Mill and Marx, and gives a spirited and controversial answer to the question underlying the whole discussion: why did Mill and Marx get it wrong?

208 pages, **£15.00** (0 631 13691 6)

Truth, Knowledge and Modality

Philosophical Papers, Volume 3

G.H. von WRIGHT

In this third volume, Professor von Wright concentrates on the ideas of truth, knowledge and modality, and the articles included are possibly more original and more recent than those of the earlier volumes. In subject matter, he is closer here to the topics which concern him most. and which will be of greatest interest to other philosophers.

176 pages, **£15.00** (0 631 13367 4)

The Legacy of Wittgenstein

ANTHONY KENNY

In these marvellously clear essays, Anthony Kenny provides an extensive commentary on Wittgenstein's work. Going beyond exegesis, he also gives a critique of other philosophers and theorists — Chomsky, Gregory, Anscombe, Teilhard de Chardin, for example — in the light of Wittgenstein's discoveries. A general introduction assesses the contemporary relevance of his thinking.

176 pages, **£12.50** (0 631 13705 X)

Wittgenstein on Meaning

COLIN McGINN

This book puts forward an original interpretation of the *Philosophical Investigations* and related works by Wittgenstein. It examines in detail what he actually wrote about meaning, understanding, rules and private language. Wittgenstein, it is argued, was opposed primarily to the 'inner process' conception of understanding, and his alternative view — that following a rule is a technique or custom — is elucidated.

160 pages, **£12.50** (0 631 13764 5)

Basil Blackwell

Philosophy Books Wanted

PHILIP MORRIS

bookseller

Academic & rare books on all branches of philosophy purchased

31 Highgate,
Beverley,
N. Humberside.
HU17 0DN
Great Britain

Telephone: 0482 - 869453

The Journal of Philosophy

Subscriptions at $15.00/year; $12.00 to students
$20.00/year to libraries

Students' price extended to retired and unemployed philosophers. Published bimonthly 1904 to 1976, monthly thereafter. Complete volumes and all separate issues available back to January 7, 1904 (volume 1, number 1); prices as follows:

BACK VOLUMES (from the first to the penultimate year)
32-page issues, $1.50; 64-page, $3.00; special issues, $4.00.
Volumes (unbound, volume indexes included), $30.00 each.

CURRENT VOLUMES (this year and last)
64-page issues, $2.00; special issues, 96-page or over, $3.00.
Volumes (unbound, volume indexes included), $25.00 each.

ALSO AVAILABLE
Cumulative Fifty-year Index, 1904–1953; articles classified by subject and author; 452 p.; cloth, $12.00. Ten-year Supplement, 1954–1963: 98 p.; cloth $3.00, paper $2.00.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

ANALYSIS 45.1 JANUARY 1985

SUBSTITUTIONAL QUANTIFICATION AND EXISTENCE

By B. J. Copeland

THE so-called Barcan Formula, $\Diamond\exists x\phi x \mathrel{⥽} \exists x\Diamond\phi x$, was used as an axiom by Ruth Barcan Marcus in her pioneer system of quantified modal logic [1]. In response to criticisms of the Formula advanced by a number of writers, Marcus has maintained that her axiom is unproblematic provided the existential quantifier is interpreted substitutionally: it is only when the quantifier is taken referentially that the axiom appears paradoxical [2]. In [4] I presented a semantics which shows that Marcus's claim is false: the semantics treats the quantifiers substitutionally, yet the Barcan Formula is invalid (throughout [4] and this paper I am concerned exclusively with an S5 setting). Since the publication of [4] I have several times encountered the objection (notably from Barcan Marcus herself, in correspondence) that my argument there fails since the semantics does not give a purely substitutional interpretation of the quantifiers, but ties quantification to existence, thus smuggling in the same 'referential bias' which pervaded earlier criticisms of the Formula. In this note I will defend the argument of [4] against this objection. En route I will respond to van Inwagen's doubts about the intelligibility of substitutional quantification [9].

The purely substitutional interpretation of the existential quantifier is this: $\exists xA$ is true (in world w) if and only if there is a member t of the substitution class such that A' is true (in w), where A' is the result of replacing any free occurrences of x in A by t, and the substitution class is some pre-selected class of expressions of the language (compare Kripke [7], pp. 329–30). For the present purposes attention may be restricted to substitution classes all of whose members are singular terms. Let me restate the semantics presented in [4] in a way which makes it clear that the treatment given to the existential quantifier is purely substitutional. A *model* is an ordered triple $\langle \mathrm{K}, \mathrm{Q}, \mathrm{v}\rangle$, where K is a set, Q is a function from K to the set of sets of singular terms of the language, and v is a function from $\{\text{formulae}\}\times \mathrm{K}$ to the set $\{T, F\}$. (The idea, of course, is that $\mathrm{v}(A, w) = T$ should be understood as asserting that A is true in possible world w.) Given K and Q, v is defined by the following recursion. For all $w \in \mathrm{K}$, all variables x of the language, all formulae A, B: $\mathrm{v}(A, w)$ is arbitrary if A is atomic; $\mathrm{v}(\sim A, w) = T$ iff $\mathrm{v}(A, w) = F$; $\mathrm{v}(A \supset B, w) = T$ iff either $\mathrm{v}(A, w) = F$ or $\mathrm{v}(B, w) =$

T; $v(\Box A, w) = T$ iff for every $w' \in K$ $v(A, w') = T$; $v(\exists xA, w) = T$ iff for some term t in $Q(w)$ $v(A', w) = T$, where A' is the result of replacing any free occurrences of x in A by t. Derivatively: $v((x)A, w) = T$ iff for every term in $Q(w)$ $v(A', w) = T$; $v(\Diamond A, w) = T$ iff for some $w' \in K$ $v(A, w') = T$; $v(A \mathrel{-\!\!\!3} B, w) = T$ iff for every $w' \in K$ either $v(A, w') = F$ or $v(B, w') = T$. A formula A is true in a model $\langle K, Q, v\rangle$ iff $v(A, w) = T$ for every $w \in K$; and a formula is valid iff true in all models.

The Barcan Formula is invalid in this semantics (details of the proof are given in [4]). To secure the validity of the formula, it must be postulated that for every model $\langle K, Q, v\rangle$ and for every w, $w' \in K$, $Q(w) = Q(w')$. Merely giving the purely substitutional interpretation of the existential quantifier is, then, insufficient to validate the Barcan Formula: one needs in addition the postulate that membership of the substitution class is world independent. (One way of building this postulate into the semantics is simply to suppress the function Q and select some fixed set as the substitution class, for example the set of *all* singular terms of the language.) Thus there is no special connection between the validity of the Barcan Formula and the use of substitutional quantification. When the quantifiers are interpreted referentially, one needs a special postulate concerning constancy of domain of quantification if the Barcan Formula is to become valid (see Kripke [6]): when the quantifiers are interpreted substitutionally, one needs a special postulate concerning constancy of the substitution class if the Barcan Formula is to become valid.

* * * * *

An uninterpreted formal semantics is no more than a mathematical characterisation of the valid sentences of the logic in question: it is only once the symbolism of the semantics has been explained in terms of antecedently grasped concepts bearing on the use of language that the semantics is capable of assigning *meanings* to expressions of the language under consideration (see my [3] and [5] for further remarks on this issue). Thus for example the above semantics assigns no meaning to $\Box$ until it is explained that T and F are truth values, that $v(A, w) = T$ (alternatively, F) asserts that A is true (alternatively, false) in possible world w, and that, in each model, K represents the set of all possible worlds. (This point has been made by Plantinga [8], p. 127.) As it stands, then, the above semantics assigns no determinate meaning to the existential quantifier, since the formal notion of a substitution class, $Q(w)$, is as yet uninterpreted. It is only once an explanation has been given of what is signified by membership of the substitution class that the semantics can properly be said to assign meanings to the quantifiers; and differing ways of interpreting the notion of a substitution class will lead to differing assignments of meaning to the quantifiers.

Van Inwagen [9] has questioned the intelligibility of substitutional quantification on the ground that 'friends of substitutional quantification' (as he calls us) offer *some substitution instance of A is true* as the truth condition for $\exists xA$, but deny that $\exists xA$ *means* that some substitution instance of A is true. This is true but, viewed aright, no cause for concern. As I have said, it is only once the formal notion of a substitution class (which is, of course, presupposed by this truth condition) has been interpreted that the truth condition will assign a determinate meaning to the existential quantifier. Moreover, what meaning is assigned will depend on what interpretation is given.

In [4] I suggested that $Q(w)$ be understood as the set of terms which *denote* in world w, and it is presumably this suggestion which has engendered the charge of referential bias which I mentioned. The point I want to stress is that the argument of [4] in no way depends upon $Q(w)$ being understood in this way: no matter *how* $Q(w)$ is interpreted it remains the case that the Barcan Formula is invalid in the semantics, and is rendered valid only if it is stipulated that the substitution class is world invariant.

The foregoing considerations prompt two questions.

(1) Is there an interpretation of the notion of a substitution class relative to which it is natural to stipulate that the function Q be single-valued?

(2) Is there an interpretation of the notion of a substitution class relative to which it is natural to stipulate that Q is *not* single-valued but which (unlike the interpretation suggested above) does not involve tying quantification to existence?

I leave (1) as an open question. An affirmative answer to it would show that there *is* an assignment of meaning to $\exists$ under which the Barcan Formula is unproblematic. An affirmative answer to (2), however, will establish that it is not only 'existence readings' of $\exists$ which make the Barcan Formula problematic. I shall proceed to show that this is indeed the case.

Let me use 'denotation' in a Meinongian way: the denotation in world w of term t may or may not actually exist in w. Call t a *denoting term* for world w if and only if the denotation of t (in w) is an object which actually exists in w. Call t a *lore term* for world w if and only if the denotation of t (in w) features in a myth, tale, tradition or theory of the inhabitants of w. A lore term for w may or may not also be a denoting term for w. (A distinction between denoting terms and lore terms can of course be drawn without using a Meinongian framework. For example, t is a lore term for world w if and only if t, or some term having the same sense as t, occurs in the verbal presentation of a myth, tale, tradition or theory of the inhabitants of w.) $Q(w)$ is now interpreted as being the union of the denoting terms and the lore terms for w. Given this way of understanding the notion of a substitution class, it is natural to allow Q to take different values at different worlds. The term 'Pegasus', for example, is a lore term for the actual world, and hence is a member of the substitution class for the actual world; but it is easy to imagine worlds for which 'Pegasus' is neither a lore term nor a denoting term.

The University of Canterbury,
Christchurch, New Zealand

© B. J. COPELAND 1985

REFERENCES

[1] Barcan Marcus, R. 'A functional calculus of the first order based on strict implication', *Journal of Symbolic Logic*, 11 (1946), 1-16.
[2] Barcan Marcus, R. 'Interpreting quantification', *Inquiry*, 5 (1962), 252-9.
[3] Copeland, B. J. 'On when a semantics is not a semantics', *Journal of Philosophical Logic*, 8 (1979), 399-413.
[4] Copeland, B. J. 'A note on the Barcan formula and substitutional quantification', *Logique et Analyse*, 97 (1982), 83-6.
[5] Copeland, B. J. 'Pure semantics and applied semantics', *Topoi*, 2 (1983), 197-204.
[6] Kripke, S. 'Semantical considerations on modal logic', *Acta Philosophica Fennica*, Fasc. 16 (1963), 83-94.
[7] Kripke, S. 'Is there a problem about substitutional quantification?' In *Truth and Meaning*, ed. Evans, G. and McDowell, J., Oxford, 1976.
[8] Plantinga, A. *The Nature of Necessity*, Oxford 1974.
[9] van Inwagen, P. 'Why I don't understand substitutional quantification', *Philosophical Studies*, 39 (1981), 281-5.

WIGGINS, ARTEFACT IDENTITY AND 'BEST CANDIDATE' THEORIES

By H. W. NOONAN

A PLAUSIBLE principle governing identity is that whether a later individual, y, is identical with an earlier individual, x, cannot ever merely depend on whether there are, at the later time, any better candidates than y around for identity with x. I shall call this principle 'the only x and y principle'. It has been a bone of contention amongst philosophers interested in identity for many years. In *Sameness and Substance* (Oxford, 1980) David Wiggins, who has always been one of the stoutest defenders of the only x and y principle, includes a formulation of it among his D theses as thesis D(x) and presents what he believes to be 'the most fundamental objection' to the opposing 'best candidate' approach to identity, i.e. the approach involving rejection of the only x and y principle.[1] In what follows I argue that Wiggins' objection fails, but that there is another objection to the 'best candidate' approach which succeeds.

An example which makes rejection of the only x and y principle seem plausible, and also brings out exactly what that involves, is that of the Ship of Theseus, as described by Thomas Hobbes (*De Corpore*, 11 (Molesworth, p. 136)). And it is in connection with this example, also, that Wiggins presents his fundamental objection to the 'best candidate' approach to identity.

In this case it does seem plausible to say that the continuously repaired ship is the Ship of Theseus but that the plank-hoarder's ship is a candidate too, and *would* have been the Ship of Theseus if its rival, the continuously repaired ship, had not existed. To insist that the plank-hoarder's ship is not the Ship of Theseus even when its rival does not exist seems absurd – for ships can certainly undergo dismantling and reconstruction; while to insist that it is still the Ship of Theseus when its rival does exist seems implausible. But even if a case could be made out for saying the latter (e.g. by reference to the antiquarian interest in the Ship of Theseus), it is clear that any considerations in favour of the plank-hoarder's ship can lend it plausibility at best, so that even if it is not definitely true that the continuously repaired ship is the Ship of Theseus in the situation Hobbes describes, it is not definitely false either. The identity statement in question is at worst indeterminate in truth value.

Moreover, even if he could produce a conclusive argument that the plank-hoarder's ship was the Ship of Theseus both in the situation in which it alone exists and in that in which its continuously repaired rival exists too, the defender of the only x and y principle would still be in apparent danger of refutation by the puzzle of the Ship of Theseus. For that would be compatible with the plank-hoarder's ship merely being the *strongest candidate* for identity with the Ship of Theseus, and the defender of the only x and y principle would still appear to have a problem on his hands so long as it seemed plausible that the continuously repaired ship was the Ship of Theseus in the absence of the plank-hoarder's ship.

So much, then, for the way in which the example of the Ship of Theseus lends plausibility to the 'best candidate' approach to identity. Now, however, we need to look at the consequences of rejecting the only x and y principle to deal with the puzzle.

We can picture the three relevant situations as overleaf.

Here the second situation is the one described by Hobbes. The first situation is that in which the Ship of Theseus undergoes repair and replacement of parts, but the replaced planks are simply discarded or destroyed, and the third situation is that in which no repair or replacement work takes place, but the Ship of Theseus is dismantled plank by plank and later reconstructed.

---

[1] Wiggins' D(x) is the following: 'what we need, if identity is what we want to elucidate, is a criterion which will stipulate that for a relation R to be constitutive of the identity of *a* and *b*, *a*'s having R to *b* must be such that objects distinct from *a* or *b* are irrelevant to whether *a* has R to *b*'. He also calls it 'the only *a* and *b* rule'.

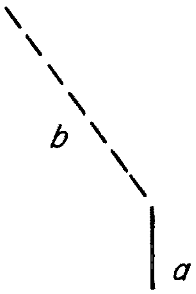

Figure 1

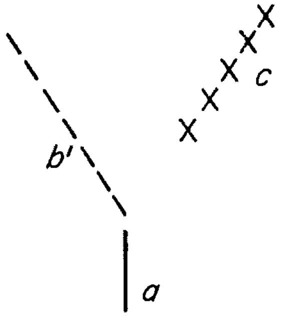

Figure 2

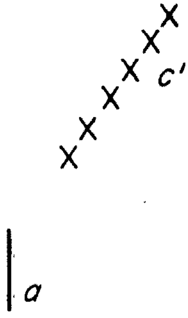

Figure 3

In each figure the continuous line represents the history of the Ship of Theseus before repair and replacement work begins. The line of dashes in Figures 1 and 2 represents the history of the continuously repaired ship, and the line of crosses in Figures 2 and 3 represents the history of the ship reconstituted from the original planks of the Ship of Theseus, i.e. the plank-hoarder's ship.

If we rigidly designate the ship originally referred to in all three situations as 'the Ship of Theseus' by 'a', the ship undergoing continuous repair in situation 1 by 'b', the ship undergoing continuous repair in situation 2 by 'b'', the ship reconstituted from the original planks in situation 2 by 'c', and the ship reconstituted from the original planks in situation 3 by 'c'', then if we reject the only x and y principle we can, in conformity with the plausible view that both the continuously repaired ship and the plank-hoarder's ship are candidates for identity with the Ship of Theseus, though the former has the stronger claim, assert that in situation 1, $a = b$, in situation 2, $a = b'$ and $a \neq c$, and in situation 3, $a = c'$. These assertions are compatible with the necessity of identity if $c \neq c'$, otherwise not. But is a 'best candidate' theorist committed to holding that $c = c'$? Wiggins assumes that he is, and this is the basis of his objection to the 'best candidate' approach. He writes:

> The most fundamental objection to the 'best candidate' approach is that it licenses the following as a possibility: we could walk up to the antiquarian's ship [i.e. c], seen as a candidate to be Theseus' ship, and say that, but for the existence of its rival, i.e. the distinct coincidence candidate that is the constantly maintained working ship plying once yearly to Delos, it would have veritably coincided as a ship with Theseus' original ship. But the idea that in *that* case it would have been Theseus' very ship seems to be absurd. ... We are discovering in this argument ... the real intuitive grounds for doubting that anything might have been a numerically different entity from the one it actually is (*Sameness and Substance*, p. 59).

But in fact the 'best candidate' theorist need not accept that $c = c'$. Of course, he has to accept this if he takes the candidates he is concerned with to be the two later *ships*, and takes it that what they are candidates for is literally *identity* with the (one and only) original Ship of Theseus. But he need not take this view. If, for example, like Robert Nozick, he regards continuants as 'four dimensional worms', composed of temporal parts or stages, he can take the competing candidates to be not ships but *ship-stages*. Then he can say that what they are candidates for is *not* identity with the original Ship of Theseus, but rather: *being at the later time the stage (temporal part) of the Ship of Theseus existing then*. The ship-stage present in situation 3 at the location of ship c' possesses this property there, but only contingently if in situation 2 $a = b'$ and $a \neq c$, for then, while still existing in situation 2, it does not possess it there. But it does not follow that there is any *ship* which exists both in situation 2 and in situation 3 and is identical with the Ship of Theseus in situation 3 but distinct from it in situation 2. For despite appearances the ship c, not being identical with the ship c', is not present in situation 3 at all (see R. Nozick *Philosophical Explanations*, Oxford, 1981, p. 656 ff). Nor does the best candidate theorist need to be a temporal worm metaphysician in order to deny the identification of c with c'. Consistently with the rejection of a temporal worm metaphysics he can, for example, take Nathan Salmon's line, deny that $c = c'$ and take the rival candidates involved in the case to be, neither ships nor ship-stages, but rather the *hunks of matter* constituting b' and c. Then he can say that what these are rival candidates for is, again, not identity with the original Ship of Theseus, but rather: *being at the later time the hunk of matter constituting the Ship of Theseus then*. The hunk of matter constituting c' in situation 3 possesses this property there, but only contingently if in situation 2 $a = b'$ and $a \neq c$, since then, though it certainly exists in situation 2 (where it constitutes c) it does not possess it there. But it does not follow that there is any *ship* in situation 3 which is identical with the one and only Ship of Theseus there, but distinct from it in situation 2, for this contingency of constitution is distinct from the contingency of identity and does not entail it, and despite appearances ship c is not present in situation 3 at all (see N. Salmon *Reference and Essence*, Basil Blackwell, Oxford, 1982, Appendix 1).

Wiggins is in error, then, in assuming that it is essential to the 'best candidate' theorist's position that he identify c with c' and thus abandon the necessity of identity. The 'best candidate' theorist has, indeed, to choose between rejecting that identification and abandoning the necessity of identity, but both courses are open to him, and, as we have seen, at least two actual 'best candidate' theorists have opted for the former (after writing the above, in fact, I came across R. M. Gale's paper 'Wiggins' Thesis D(x)', *Philosophical Studies*, 1984, in which the 'best candidate' theorist is defended against Wiggins' argument precisely on the ground that the former option is the one he should adopt). Actually, however, this is merely to jump out of the frying pan into the fire. For the objection to the 'best candidate' theorist's taking this line should be obvious. We can put it in a form reminiscent of Wiggins' objection. I have stressed that if the 'best candidate' theorist denies that $c = c'$ he must acknowledge that, despite appearances, c is not present in situation 3 at all. He must then acknowledge that the following is a possibility: we could walk up to the plank-hoarder's ship in situation 2, seen as a candidate to be Theseus' ship, and say (truly) that, but for the existence of its rival, it would never have existed *at all*. Similarly, if he takes the same line in the case of personal identity, then, if he accepts a mentalistic criterion of personal identity which entails that Bernard Williams's Charles *is* Guy Fawkes in the absence of a rival claimant, the 'best candidate' theorist is not committed, as Wiggins suggests (*Sameness and Substance*, p. 208), to the view that it is true of Charles in the reduplication situation that but for the existence of Robert *he would have been* Guy Fawkes. But to avoid this commitment he must acknowledge that we could walk up to Charles in the reduplication situation and, speaking of Robert, say to him 'You should consider yourself fortunate that that other fellow seems to be as good as you are at reminiscing about attempts to blow up the Palace of Westminster – if he hadn't been you wouldn't have existed'. But I take it that it goes without saying that in making these acknowledgements the 'best candidate' theorist would be committing himself to self-evident absurdities. (In case it is thought that the absurdity in the latter case is not one that derives from the 'best candidate' theory as such, but merely from the mentalistic theory of personal identity that licenses the identity of (the non-duplicated) Charles with Guy Fawkes in the absence of any possible causal connection between them, let me refer as well to the split brain transplant case familiar from the writings of Wiggins and Parfit. If the 'best candidate' theorist adopts the line of reply to Wiggins presently being considered then he must acknowledge that in this case one could say to either one of the fission products: 'You should consider yourself fortunate that the other fellow's brain transplant went so well – if it hadn't you would never have existed.' The absurdity is the same.)

This, then, is the sort of absurdity the 'best candidate' theorist is committed to if he denies the identification of c with c', while if he endorses that identification, of course, Wiggins' objection applies. Perhaps then I should not have said that Wiggins' objection fails, but merely that it is incomplete; that is, that he presents only one horn of the dilemma which in fact refutes the 'best candidate' theory. But either way the 'best candidate' theory *is* refuted and the only x and y principle vindicated.

University of Birmingham,
P.O. Box 363, Birmingham B15 2TT

© H. W. NOONAN 1985

THE PARADOX OF THE LIAR – A CASE OF MISTAKEN IDENTITY

By LAURENCE GOLDSTEIN

CONSIDER the (strengthened) Liar generator k, where

$$k = \text{‘}k \text{ is not true’}$$

On each side of this identity we have a different name of a putative statement. Predicating truth of this putative statement, we obtain

$$k \text{ is true}$$

or equivalently, using the other name of k,

$$\text{‘}k \text{ is not true’ is true.}$$

Thus

(a) k is true ≡ ‘k is not true’ is true

Adding ‘k is not true’ to each side of this equivalence, we obtain

(b) k is true & k is not true ≡ ‘k is not true’ is true & k is not true

Hence, using classical principles to simplify the right hand side,

(c) k is true & k is not true ≡ k is not true

But the right hand side of equivalence (c) is just the Liar generator k which is thus shown to be logically equivalent to a contradiction.

This may seem surprising, since k has the surface appearance of being the negation of an atomic statement. How can it be equivalent to a contradiction which is *conjunctive* in form? The inner complexity of the Liar generator becomes apparent with the realization that somebody asserting it would be asserting k *and* (or ‘and thereby’) asserting that k is not true (for that's what k ‘says’).

The conclusion that the Liar sentence k is a contradiction does not, of course, solve that version of the paradox. For the following seems a natural line of thought.

(d) If k is a contradiction then it is a statement that is not true.
(e) But k says just that – that it is not true.
(f) Hence k is true.

We have established the antecedent of (d), but the consequent of (d) paradoxically contradicts (f).

There are several ways of blocking this reasoning. First, and most drastically, one could reject (d) by accepting the Hegelian view that there are some *true* contradictions. Pierre d'Ailly, the late mediaeval author, maintains that the Liar sentence is a true contradiction [4] and this position is shared by some modern writers (e.g. [13]). Second, one could reject (d) by saying that contradictions are not statements and hence are neither true nor false. This account of contradictions was championed by Wittgenstein for a considerable time (e.g. [19] pp. 53-9, [20] 4.416, 5.143, [18] p. 131). Third, one could claim that the predicate 'true' is not univocal, or is incomplete, or needs to be relatavised to occasion of utterance so that it does not retain a stable character throughout the above reasoning [16], [12], [8], [9], [1]. A fourth way to block the reasoning is to reject (e) on the grounds that *k* does not *just* say that it is not true, for it *also* says that it *is* true. Buridan [10] was one of several mediaeval authors who took this line, and it has recently been revived by James Cargile [2].

Each of these proposals (and many others) designed to stem the paradox reasoning have been effectively criticized and all of them lack the intuitive appeal of the few principles utilized in the argument (a)-(f). Let us then take seriously the idea that this simple argument is, consistent with appearances, *valid*, and that the premises (d) and (e) are true. If a valid argument leads to a contradiction then, by *reductio*, a premise may be validly rejected. Thus, by elimination, it is the original identity claim '*k* = "*k* is not true" ' that is not true.

This conclusion may at first appear absurd because, so it may be objected, the original identity claim results from a *stipulation* – we construct a statement '*k* is not true' and decide to call it *k* – the name is chosen at pleasure without any restriction. The objection is, however, mistaken. A shorter proof that there is some restriction runs as follows: *If*, in the identity claim

$$k = \text{`}k \text{ is not true'},$$

the statement named on the left were identical with the statement named on the right, then the statement named on the left would be true if and only if the statement on the right were true, i.e.

$$k \text{ is true} \equiv \text{`}k \text{ is not true' is true}$$

So

$$k \text{ is true} \equiv k \text{ is not true}$$

But this is a contradiction, hence the original identity claim must be rejected.

Another way of seeing the plausibility of this restriction is to start by considering the biconditional

$$k \text{ is true} \equiv k \text{ is not true}$$

One could reason: if *k* is true, then it is not, but if it is not, then it is. So if we suppose *j* to be the name of the biconditional, it is clear that *j* is not true because *there can be no statement k* that satisfies that pair of conditionals. Can we now go self-referential by letting '*k*' be a name of that biconditional? It may seem correct to say that since *j* was shown to be an untrue statement, the same must hold of *k*. Yet this is not so, because '*j*' was a name chosen at pleasure; the name *k*, as we have just seen, names no statement so we are *not* free now to use it as a name of an (untrue) statement. Hence we cannot identify *k* and *j*.

According to the Buridan-Cargile approach mentioned earlier, sentences generating versions of the Liar paradox are contradictions and hence false. But, if our argument is correct, it is only by ignoring a condition on identity that Liar generators can seem to be contradictions. A suggestion that has been pursued by Brian Skyrms [14], [15] is that the law of substitutivity of identicals, when used in the usual Liar reasoning, fails. Our conclusion is that it is not failure of this law but a failure to substitute identicals that gives rise to the contradiction. Interestingly, paradoxes in the family studied by Thompson [17], including the Barber, Russell's and Grelling's, can also be shown to rest on contradictory assumptions resulting from the breach of an identity condition – see [3] and especially [6] §2.3.

Exactly what sin is committed by somebody who utters a Liar generator? Are we obliged to allow that what he says is true or not true and, if so, does not paradox ensue? Consider a person, X, who makes the statement '*k* is not true'. Here '*k*' names a statement said by X to be not true. So X is asserting a statement that predicates non-truth of *k*. Hence it is not for X simply to 'let' '*k*' be the name of *that* statement, for he has *already* assigned the name '*k*' to a statement he has said is *not* true, namely *k*. Even if X were lying and, despite what he says, believes *k* to be true, he still cannot believe the contradictory statement '*k* is not true' to be true, so *a fortiori* he cannot believe it to be the same statement as *k*. X may or may not be lying, but he cannot consistently believe that *k* is identical with the statement that *k* is not true. This is not to attribute an unrealistically high level of consistency to X. Rational agents may sometimes be committed to inconsistent beliefs, but they do not believe in contradictions. Of course, it may be the case that X is using the name '*k*' ambiguously as between 'k_1' and 'k_2'. But no paradox arises from the identity claim '$k_1 =$ "k_2 is not true"' when $k_1 \neq k_2$. '*k*' cannot name the very statement in which non-truth is predicated of *k*. The following comparison may be helpful. In the programming language BASIC, the sentence '$N = N + 1$', which may be read 'let N have a value one greater than its previous value' is used to assign N a new value while *cancelling* the old value. But to 'let' *k* be '*k* is not true' is, *per impossibile*, to let one statement, whatever its truth-value, *also* have a *different* truth-value.

The restriction on designation-assignment that we have discovered arises only out of a simple consistency requirement: if '*a*' designates a statement and '*b*' designates a negation of that statement, then $a \neq b$. It is easy for a theoretician to overlook this restriction since, on the classical (Tarskian) theory of quotation, '*k*' as it occurs inside the quotation-expression '*k* is not true' is deemed just a typographical part of a *name* – it has no independent semantical role – so we appear free to assign '*k*', when it appears outside quotation marks, the semantical role of designating anything we please. However, the Tarskian theory of quotation is false (see [5], [7]). Quotation marks do not serve to form semantically unstructured names, but are *pointing* devices, and, in the quotation-expression '*k* is not true' as we have been using it, '*k*' is pointed to as the designator of a statement of which non-truth is predicated. Hence the restriction on using '*k*' for some incompatible purpose. Anyone unconvinced about the necessity for 'reading inside' quotation marks should consider whether he would accept the putative identity stipulation

$$3 = \text{What '}3 + 1\text{' designates.}$$

One point to be noted about the restriction on designation-assignment for which we have argued above is that it does not entail a general ban on self-reference. Thus nothing we have said would count against the possibility, mentioned by Kripke [12], of using 'Jack' to name the sentence 'Jack is short'. Nor have we said anything to suggest that true statements such as 'This statement is self-referential', 'This statement is quite interesting' and 'This statement is made by the use of an English sentence' are somehow illegitimate.

A semantical restriction that may be difficult for the theoretician to uncover may nonetheless be such that any competent speaker would conform to it and recognize when it is breached (compare our awareness of syntactical transgressions). No competent speaker would seriously assert that the very statement he is making is not true. This is an important, but generally neglected point about the Liar. Liar-type sentences may be used in jest, or for some other non-truth-claiming purpose, but if someone seriously wished to *assert* one of them we should have to declare him semantically incompetent since he fails to observe a fundamental condition on identity or, correlatively, a fundamental restriction on designation-assignment. Aristotle, in [11], urges the similar point that someone who does not conform to the principle of non-contradiction cannot talk about anything at all. A person who purports to make a candid assertion of a Liar sentence does not share our ways of thought and talk. As Wittgenstein in another context puts it, 'He is incomprehensible to me means that I cannot relate to him as to others' [21] §198. We can attach no sense to what the person who flouts our semantical principles says; he uses our words but does not use them our way. What he utters – his sentences – may be in perfect grammatical order, but *what he means* is inaccessible to us.

Hong Kong University,
Pokfulam Road,
Hong Kong

© Laurence Goldstein 1985

REFERENCES

[1] T. Burge, 'The Liar Paradox: Tangles and Chains', *Philosophical Studies*, 41 (1982), 353-66.
[2] J. Cargile, *Paradoxes: A Study in Form and Predication*, Cambridge: Cambridge University Press, 1979.
[3] C. Chihara, 'The Semantic Paradoxes: A Diagnostic Investigation', *The Philosophical Review*, 88 (1979), 590-618.
[4] P. d'Ailly, *Concepts and Insolubles*, Dordrecht: Reidel, 1980.
[5] D. Davidson, 'Quotation , *Theory and Decision*, 11 (1979), 27-40.
[6] L. Goddard and M. Johnston, 'The Nature of Reflexive Paradoxes: Part I', *Notre Dame Journal of Formal Logic*, 24 (1983), 491-508.
[7] L. Goldstein, 'Quotation of Types and other Types of Quotation', *Analysis*, 44.1 (January 1984), 1-6.
[8] A. Gupta, 'Truth and Paradox', *Journal of Philosophical Logic*, 11 (1982), 1-60.
[9] H. Herzberger, 'Notes on Naive Semantics', *Journal of Philosophical Logic*, 11 (1982), 61-102.
[10] G. E. Hughes, *John Buridan on Self-Reference*, Cambridge: Cambridge University Press, 1982.
[11] C. Kirwan, *Aristotle's Metaphysics Books* Γ, Δ, *E*, Oxford: Clarendon Press, 1971.
[12] S. Kripke, 'Outline of a Theory of Truth', *Journal of Philosophy*, 72 (1975), 690-716.
[13] G. Priest, 'The Logic of Paradox', *Journal of Philosophical Logic*, 8 (1979), 219-41.
[14] B. Skyrms, 'Notes on Quantification and Self-Reference', in R. L. Martin (ed.), *The Paradox of The Liar* (Second Edition), Reseda: Ridgeview, 1978.
[15] B. Skyrms, 'Intensional Aspects of Semantical Self-Reference' in R. L. Martin (ed;), *New Essays on the Liar Paradox*, Oxford: Oxford University Press, 1984.
[16] A. Tarski, 'The Concept of Truth in Formalised Languages' in his *Logic, Semantics and Metamathematics*, Oxford: Clarendon Press, 1956.
[17] J. F. Thomson, 'On Some Paradoxes' in R. J. Butler (ed.), *Analytical Philosophy* (First Series), Oxford: Basil Blackwell, 1962.
[18] F. Waismann, *Ludwig Wittgenstein and the Vienna Circle*, Oxford: Basil Blackwell, 1979.
[19] L. Wittgenstein, *Notebooks 1914-1916*, Oxford: Basil Blackwell, 1961.
[20] L. Wittgenstein, *Tractatus Logico-Philosophicus*, London: Routledge Kegan Paul, 1922.
[21] L. Wittgenstein, *Last Writings on the Philosophy of Psychology*, Oxford: Basil Blackwell, 1982.

SET THEORY, SKOLEM'S PARADOX AND THE *TRACTATUS*

By A. W. MOORE

'We say: but that *isn't* how it is! – it *is* like that though! and all we can do is keep repeating these antitheses.' (Ludwig Wittgenstein)

WITTGENSTEIN'S *Tractatus* is characterized, famously, by the following predicament: it is, by its own lights, nonsense. In it, Wittgenstein attempts to talk about the world and its limits (the limits of what is and what can be represented), these not being themselves anything *in* the world. But 'the world' is supposed to be understood schematically as containing everything that can be meaningfully spoken of. So there is no legitimate way of talking about its limits, nor indeed about *it*. If we attempt to do so, we shall produce either arrant nonsense or claims which, though intelligible, have to be understood in the wrong way; they will have to be understood as claims about items within the world, and, in purporting to concern something other than those items, they will inevitably include downright falsehoods.

Something akin to this predicament confronts anybody with a certain view of what sets are like, a view which I shall outline shortly. It is itself a predicament with a variety of aspects, and I shall later argue that the only real paradox to which the Löwenheim–Skolem Theorem gives rise (which is what I, perhaps unusually, shall mean by Skolem's Paradox) is best viewed as one of these aspects. Since the underlying view of what sets are like has great intuitive appeal, Skolem's Paradox will thus appear as a genuine paradox. That there is such a paradox connected with the Löwenheim-Skolem Theorem is evidenced by the sense of irresoluble conflict which one is liable to feel when reading the extensive literature in this area. (See especially the debate between Resnik and Thomas in [5], [6], [8] and [9].[1]) Of course, a linchpin of the *Tractatus*, inextricably linked to the attendant predicament, is Wittgenstein's doctrine that what he attempts to say in this work, even though it cannot be said, nevertheless manifests itself in (or is shown by) what *can* be said. I shall suggest that this too has its echoes in the set-theoretical case, and that recognizing these echoes may help to explain why at least the main debate occasioned by Skolem's Paradox, namely the debate between relativists and non-relativists, is destined to remain irresoluble.

The underlying view of what sets are like is the one which motivates most contemporary work in set theory (cf. [1]). The term 'set' is to be understood as a technical term, in such a way that the members of any set must themselves be sets. Still there is the Null Set, the set to which no set belongs; then there is the set whose only member is the Null Set; then these two themselves combine together to form a third set; and in this way infinitely many sets can be accounted for. But there exist infinite sets too, subsets of this infinite stock which do not themselves already appear there. These in turn belong to further sets, and so on forever. According to this picture, sets form a 'V'-shaped, hierarchical structure. The Null Set lies at its base and every other set lies somewhere further up, constituting a collection of sets taken from below it.

Although the term 'set' is being used technically, this account is intended to capture the rough intuition that any collection of sets will in turn be a set. No distinction is being recognized between sets and proper classes (that is to say, between those collections of sets which may, and those which may not, themselves be deemed sets). Provided that we do think of sets in the broad way suggested, as nothing other than collections of collections, the distinction is intuitively groundless. What this means, however, given that there is no end to the process of set formation, is that there can literally be no such thing as the totality of all sets (the set hierarchy). For any such totality would itself have to be a set, thus lying somewhere *within* the hierarchy and thus failing to contain every set. (It would fail to contain itself, for example.) It is a fundamental principle of this way of looking at sets that there is no set which contains every set as a member.

---

[1] Putnam's interesting work in this area (principally [4]) bears on the most fundamental issues in the philosophy of language but is not directly related to what I am calling Skolem's Paradox. His chief concern is with a very general problem about how far linguistic practice can determine reference whereas Skolem's Paradox, on my view, is a paradox specifically about *sets*.

This is already a paradox (cf. [3], pp. 86–7). Certainly, we have a pretheoretical intuition that, if we are able to make generalizations about things of a certain kind, then those things can be grouped together into a determinate totality. Yet here we are making generalizations about sets, at the same time having to admit that there is no totality of all sets. (The Burali-Forti Paradox is very closely related to this. There can be no totality of all ordinal numbers either, for any such totality would itself have to be an ordinal number different from, because greater than, all those belonging to it. Yet we seem to be able to make perfectly legitimate generalizations about ordinal numbers, and hence the paradox.[2]) But paradoxicality comes in degrees, and there is no reason as yet to suppose that these paradoxes cut deeper than, let us say, the paradox that simultaneity is relative to a frame of reference; that is, they may be paradoxical only to the extent that they force us to revise certain inchoate intuitions. In assimilating this picture of what sets are like, we must learn to accept that there is nothing (no thing) which contains them all: there is no set of all sets. It is nevertheless significant that as soon as we attempt not just to say things about sets but to say things *about* saying things about sets, viewing the subject matter of set theory from without, then the intuition that there must be such a thing is particularly difficult to resist. This is because the relevant notion of a thing here is simply that very general notion of whatever is picked out by a singular term (if it picks out anything) whilst the relevant notion of a singular term is in turn sufficiently general to include a whole range of phrases which, when we are attempting to give an account of the semantics of set-theoretical discourse, we find it enormously tempting to use: 'the range of the quantifiers "every set" and "some set"'; 'the relation (set of ordered pairs) picked out by the predicate "... is a member of ___"'; 'the hierarchy of sets'. These are phrases none of which, on the given understanding of what sets are like, does pick out anything, but each of which tenaciously presents itself as though it did, lest our set-theoretical discourse should fail to make sense. We are tempted to say that the subject matter of such discourse just is the hierarchy of sets. And thus we want to view the hierarchy of sets as an entity in its own right, in effect as a colossal set, which is nevertheless not itself anything *in* the hierarchy. It is impossible to talk about such an entity, however, without representing it as a common or garden set and thereby making claims which, if they are intelligible at all, have to be understood as basic set-theoretical falsehoods – for example, that there is a universal set. Of course, it is still possible to view such claims as arising from a misguided intuition which must be eradicated. But it cannot be denied that they *seem* to elucidate something which is apparent in (shown by?) the very fact that we can make generalizations about sets. I wish to argue that the more we reflect on our set-theoretical discourse, particularly from a formal point of view, the more compelling is the thought that there really is something to be elucidated in this way. The intuition which we successfully excise from what we actually say about sets is revitalized by reflections on the relations between formal languages and ways of interpreting them. This is the predicament alluded to earlier. Its structural similarity to the predicament faced by Wittgenstein in the *Tractatus* is, I hope, self-evident, as is the fact that the paradoxes mentioned above are merely aspects of it. I shall argue that Skolem's Paradox is another such.

---

[2] The concept of a set and the concept of an ordinal number are what Dummett has called indefinitely extensible concepts, and it is this feature of indefinite extensibility which lies at the root of these paradoxes. See [2], pp. 195–7.

It is generally accepted that any claim which can be made about sets and which is of mathematical significance can be expressed in a first-order language whose sole non-logical constant is the two-place predicate ϵ ('... is a member of __'). We naturally assume that any sentence from this language will be true or false under the intended interpretation, as informally described above. This assumption is not entirely innocuous. We may not be entitled uncritically to accept the principle of bivalence. Nor is it obvious that the loose description of sets offered does actually warrant our talking about *the* intended interpretation. (There are certain sentences, for example the Generalized Continuum Hypothesis and perhaps also those sentences which assert the existence of very large cardinals, concerning which our intuitions begin to run dry and whose truth-values simply may not be settled by what, according to anything that has been said so far, ϵ is supposed to mean.) We shall nevertheless allow ourselves the luxury of waiving these anxieties. There is, in the context of this discussion, a more substantial difficulty which needs to be faced.

One truth about sets, which can be established by Cantor's diagonal argument and couched in this first-order language, is that there is no one:one correlation between ω, the set of natural numbers, and its own power-set; that is, uncountably many sets belong to $P(\omega)$. Let σ be the sentence which expresses this truth.

It is a consequence of the Löwenheim–Skolem Theorem, in the version whose proof requires the Axiom of Choice, that any true sentence from this language, under the intended interpretation, will still be true under an interpretation which results from the intended interpretation by the elimination of all but countably many of its sets. In particular, this is true of σ. Not that this is paradoxical; nor does it constitute our difficulty. The reason why σ remains true under the restricted interpretation is that it now says that there is nothing acting as a one:one correlation *in the restricted domain* between what is being represented as ω and what is being represented as $P(\omega)$, this being perfectly compatible with the fact that there nevertheless does exist such a correlation. Nor is there anything to cause alarm in the bare and unsurprising fact that unintended interpretations of the specified language can differ radically from the intended interpretation, even when they confer truth on precisely the same sentences. Our difficulty does not concern points of contrast between the intended interpretation and any others; it concerns the intended interpretation itself. For simply in specifying it, and in drawing attention to the differences which there are between it and the restricted interpretation, as we must if we are to frame an appropriate response to the Löwenheim–Skolem Theorem, we face the problem adumbrated two paragraphs back; moreover, as I shall attempt to argue in due course, there is a special way in which the Löwenheim–Skolem Theorem aggravates this problem. It will not do, of course, to say that the intended interpretation has a domain of sets and that it assigns to $\in$ the set-membership relation defined over that domain. The restricted interpretation satisfies this condition. What we *want* to say is that the intended domain of quantification comprises all sets (and that the restricted domain, by contrast, fails to contain at least certain subsets of ω). But there is no such thing as the totality of all sets.

To be sure, it is still not clear that this problem presents us with a deep and ineradicable paradox. The semantics for first-order languages are usually given in such a way as to suggest that the quantifiers are to be understood as ranging over a particular collection of objects. But perhaps this suggestion can be annulled. Perhaps we can fix the intended interpretation of any sentence from the language of set theory, of the form $\forall x A$, simply by saying that it will be true just in case every set satisfies the condition specified by A – neither more nor less than that. There will then be no need to admit a totality of all sets. Nevertheless, we shall have to proceed delicately. If we are to treat the language of set theory as an object of study and discuss the intended interpretation of that language, then we must avoid the trap of doing so in such a way that we construe the component parts of the interpretation as things whose existence is ruled out on set-theoretical grounds. Nor shall we achieve this goal simply by stipulating that the intended domain of quantification, whatever else we say about it, is not a *thing*; as has already been remarked, the notion of a thing here is so broad that we had better not talk about the intended domain of quantification at all. We have thus once again been brought face to face with our predicament: various compelling ways of talking about the subject matter of set theory from without can be understood only in such a way that we are saying false things about it from within. The Löwenheim–Skolem Theorem exacerbates the predicament. It greatly enriches what we seem compelled to say, but cannot say, from without. For, to the extent that the intended interpretation of the language of set theory does seem to incorporate a domain of sets, so too it will seem that there are certain collections (for example, the domain itself) which do not belong to it, but which, from a higher vantage-point, can themselves be construed as sets belonging to a wider domain. But from that vantage-point, there may be access to correlations whose existence in the original domain some sentence of the language, so interpreted, quite correctly denies – correlations which, given the Löwenheim–Skolem Theorem, may show that sets which appear uncountable from the original vantage-point can, from the new vantage-point, be represented as countable. The only possible conclusion seems to be that notions such as countability and uncountability are inherently relative. Thus, it seems, our description of $P(\omega)$ as uncountable, even though correct, must be understood relative to our own current point of view. From another point of view, this very set may be countable.[3] But I want to argue that such relativism, compelling though it is, is subject to the by now familiar predicament that any statement of it, if it is to be intelligible at all, will have to be understood within a framework that casts it as a straightforward error. It is this which I take to be Skolem's Paradox.

The crux of the matter is this. If there is an implicit relativization in our claim that $P(\omega)$ is uncountable (the claim which is established by Cantor's argument), then it ought to be possible to make it explicit (just as it is possible to make explicit any relativization in the claim that a given physical object is moving). But it is possible to do this only insofar as it is possible to construe our discourse about sets as discourse about a particular collection of objects, the collection to which such claims must be relativized. And this in turn is not possible unless we endorse the fundamental error that there is a set which contains all the sets we intend to talk about. When it is claimed that $P(\omega)$ is not unconditionally uncountable, we have no way of understanding this except as the demonstrably false claim that it is not uncountable at all.

---

[3] The claim that a given infinite set is countable will, on this view, be no less in need of relativization. Indeed, such relativism will not be confirmed to claims about the infinite. Our specification of the size of the Null Set will have to be understood relative to our own current point of view, since the Null Set is, after all, defined to be the set to which *no* set belongs. Skolem himself, who first drew attention to these issues in [7], at the same time espoused some version of relativism. See pp. 295–6.

Admittedly, there are interpretations of the language of set theory under which all the "right" sentences come out true and in which ω is correctly represented as such even though the set represented as its power-set is countable. Any such interpretation can be thought of as a limited point of view; there are correlations to which we have access which are not in its domain. But there are also further subsets of ω to which we have access, as Cantor's argument testifies. What we ourselves take to be $P(\omega)$ never appears to be anything other than uncountable. The relativist, convinced that our own point of view is in turn limited, urges us to acknowledge the possibility that what we ourselves take to be $P(\omega)$ is not – as viewed from some even higher vantage-point, a vantage-point from which it may yet be countable. But how are we to make sense of this? Certainly not by trying to view $P(\omega)$ from two different points of view at once; that would be incoherent. Nor by trying to view it simply from this point of view; that would make the possibility unintelligible. But if it were possible to view it from an absolute standpoint, then relativism itself would lose its rationale and there could be no objection to saying that $P(\omega)$ contained all of ω's subsets and that it was unconditionally uncountable. So if we do deny the absolute uncountability of $P(\omega)$, then what exactly are we denying and where, so to speak, are we denying it? (The mere fact that there are legitimate concepts of countability and uncountability which do involve relativization to certain domains is beside the point. The relativist wants to insist that there are no *absolute* concepts of countability and uncountability – that it makes no sense to describe $P(\omega)$ as unconditionally uncountable.) We, in mounting a general investigation into what sets are like, can only aspire to know whether or not $P(\omega)$ is countable, as it were *here*. It is not. But we can have no grasp on any distinction between what is true here and what is true simpliciter. So $P(\omega)$ is uncountable simpliciter.

Yet the very use of the word 'here' appears to vindicate the relativist. There remains a real predicament. That which cannot legitimately be stated (relativism) appears, for all that, to impress itself upon us as soon as we step outside mathematical practice and relflect on what is revealed therein. This predicament is directly analogous to that which Wittgenstein faces in the *Tractatus*. There is no particular point of view *in* the world which can be spoken of as *here*: our point of view is a limit of the world. That is, there is no particular set *in* the hierarchy of sets which can be spoken of as the intended range of the quantifiers: they are intended to range over the entire hierarchy (though not even this can properly be said). And here an intriguing possibility arises. If we are prepared to extend the analogy with the *Tractatus*, then it will become apparent that, despite the fact that relativism defies any coherent statement, the debate between the relativist and the non-relativist is in a very deep sense irresoluble. For what the relativist *means* is quite correct; only it cannot be *said*, but makes itself manifest (cf. [10], 5.62).[4]

University College,
Oxford OX1 4BH

© A. W. MOORE 1985

REFERENCES

[1] George Boolos, 'The Iterative Conception of Set', in *The Journal of Philosophy*, 68 (1971).

[2] Michael Dummett, 'The Philosophical Significance of Gödel's Theorem', reprinted in *Truth and Other Enigmas* (Duckworth, 1978).

[3] Jonathan Lear, 'Sets and Semantics', in *The Journcl of Philosophy*, 74 (1977).

[4] Hilary Putnam, 'Models and Reality', in *The Journal of Symbolic Logic*, 45 (1980).

[5] Michael Resnik, 'On Skolem's Paradox', in *The Journal of Philosophy*, 63 (1966).

[6] Michael Resnik, 'More on Skolem's Paradox', in *Noûs*, 3 (1969).

[7] Thoralf Skolem, 'Some Remarks on Axiomatized Set Theory', reprinted in *From Frege to Gödel, a Source Book in Mathematical Logic, 1879–1931*, edited by Jean van Heijenoort (Harvard University Press, 1967).

[8] Williams Thomas, 'Platonism and the Skolem Paradox', in *Analysis* 28.6 (June 1968).

[9] William Thomas, 'On Behalf of the Skolemite', in *Analysis* 31.6 (June 1971).

[10] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, translated by D. F. Pears and B. F. McGuiness (Routledge and Kegan Paul, 1961).

---

[4] I thank an *Analysis* referee for comments on an earlier version.

DAVIDSON AND INDETERMINACY OF TRANSLATION

By ROBERT KIRK

DAVIDSON has always acknowledged that his approach to questions of meaning leaves open the possibility of Quinean indeterminacy of sentence translation. There would be such indeterminacy, he suggests, if two different truth theories both conformed to all the formal and empirical constraints that he recognizes (1974a, p. 321 f.; 1974b, p. 327). So much the worse, you might think, for his choice of constraints (cf. Lewis). However, in this note I want to concentrate on a different response. It is to say that if there is this sort of indeterminacy, it is negligible. If it exists at all, 'it is because when all the evidence is in, alternative ways of stating the facts remain open'; and 'indeterminacy of this kind cannot be of genuine concern'.

This response is Davidson's own (1974a, p. 322). I shall argue that it is inadequate because it either gratuitously trivializes Quine's thesis of the indeterminacy of translation, or begs the question against it.

To illuminate his conception of the possibility of indeterminacy Davidson has offered analogies with measuring subjective values and measuring temperature (1974a, pp. 314 f. and 322; 1977, p. 139). In each case numbers are assigned to represent the facts. But different assignments of numbers would serve equally well to represent the same facts. For example, where 0, 1 and 2 can be used to measure the respective values to someone of receiving £0, £4 and £11, the purposes of decision theory would have been equally well served by, e.g., 2, 4 and 6. And where 61°F serves to give the temperature here now, so does 16°C. The constraints do not determine a unique assignment of numbers, but 'the *pattern* of assignments is significant'. Analogously, he suggests, 'what is invariant as between different acceptable theories of truth is meaning'. We give the meaning of a sentence by assigning it 'a semantic location in the pattern of sentences that comprise the language'. 'Different theories of truth may assign different truth conditions to the same sentence (this is the semantic analogue of Quine's indeterminacy of translation), while the theories are (nearly enough) in agreement on the roles of the sentences in the language' (1977, p. 139).

On this account meaning (or interpretation) remains invariant between acceptable theories of truth, and sameness of meaning is sameness of semantic location in the pattern of sentences, or sameness of role in the language. However, this account of the alleged indeterminacy of translation and interpretation appears to trivialize the doctrine — gratuitously, as will become clear.

Davidson assumes that if two acceptable truth theories assign different truth conditions to a given sentence, there will be two rival, but acceptable, translation manuals which translate that sentence as differently as Quine has envisaged. That seems reasonable. But on these suggestions of Davidson's all acceptable truth theories, hence all the corresponding translation manuals, will *agree* on meaning. To any given sentence the truth theories will all assign the same semantic location, though in different ways. So there is a clear sense in which the different theories of truth and translation will not be genuine rivals. The different sets of truth conditions will just be different ways of 'stating the same facts': they will determine the same semantic location or role in the language, so that what appear to be rival translations will actually be assigned the same meaning.

When his remarks are understood in this way, Davidson is certainly entitled to say 'indeterminacy of this kind cannot be of genuine concern'. But can the implied interpretation of Quine's doctrine be right (waiving, as we must, the little difficulty that if Quine is right there is nothing to be right about)? After all, not even the most extreme advocates of objectivity and determinacy in semantics are concerned to deny that the same facts can be stated

P 4707

in different ways. They will happily accept that there are different ways of giving the truth conditions for synonymous sentences. So Davidson's interpretation of Quine's indeterminacy thesis appears to trivialize it: the thesis becomes something nobody would wish to deny. And this interpretation is quite gratuitous because it goes directly counter to Quine's explicit statements of the thesis. In one version, the sentences offered as translations of a given sentence by different manuals might 'stand to each other in no plausible sort of equivalence however loose' (1960, p. 27). In contrast, sameness of 'semantic location in the pattern of sentences' cannot be anything but a close sense of equivalence. Nor is there any help from those alternative statements of Quine's thesis in which he says for example that 'one manual would offer translations that the other would reject' (1977, p. 193). For Davidson's different truth theories actually agree on meaning, as we have seen. (His analogies with assigning numbers to subjective values and temperatures can hardly be the sort of thing Quine had in mind. No one supposes that to adopt one particular scheme of assignments of those kinds is to *reject* the ones provided by a different scheme, as Quine's doctrine requires one translator to reject the other's translation. If you give the temperature first in Centigrade and then in Fahrenheit you may confuse some people, but you're not being inconsistent.) So it seems that the trivial thesis Davidson concedes is not the one Quine is maintaining.

However, you might object that I have misconstrued Davidson's point. He is not in the business of interpreting Quine's indeterminacy thesis: he is attacking it. He knows perfectly well that Quine didn't have the trivial thesis in mind, but his point is that only that trivial thesis could be true. In fact, you may suggest, what Davidson offers us is a variation on a familiar argument against the indeterminacy thesis. This argument goes as follows. If the structure of our language, and its relations with non-linguistic reality, are such that two translations of a given foreign sentence are both equally good, then the two translating sentences cannot fail to be equivalent in a strong sense – they cannot be said to be significantly different in meaning. So it is impossible for there to be the sort of indeterminacy that Quine has envisaged. ('Structure' and 'relations with non-linguistic reality' are both to be understood, acceptably to Quine, as realized in our verbal dispositions. For versions of the argument see e.g. Dummett 1974, p. 372, and Boorse.)

It seems that for Davidson the structure is represented by the 'pattern of sentences'; and sentences which are permissibly counted as meaning the same cannot occupy significantly different slots in that pattern. However, this argument against Quine is at best incomplete. As it stands it merely begs the question against him.

Certainly there will be a considerable degree of symmetry in the dispositional structures supporting pairs of sentences that are permissibly offered by rival translation manuals as their respective versions of a given sentence. But the trouble with the argument from structure – in Davidson's version and in the others – is that Quine's indeterminacy thesis is precisely to the effect that two empirically adequate translation manuals could be in significant conflict *in spite of such symmetry*. His contention is that the facts about dispositional structure are compatible with strongly conflicting interpretations of countless sentences. To maintain or assume that this is impossible is therefore question-begging. Now you might want to define 'semantic location' and 'role in the language' in terms of, say, conventions, linguistic rules, intentions or other propositional attitudes. If you did, no doubt you would be able to rule out any possibility of conflicting schemes of interpretation or translation. But of course you would not thereby be refuting Quine, for whom such definitions are heavily contaminated by the indeterminacy (see e.g. his 1960, p. 221). Nor would it help the argument if we reverted to talk of truth conditions instead of dispositions. For it would again be question-begging to assume that whatever Davidson accepted as statements of truth conditions which went beyond hygienic statements about verbal dispositions would also be acceptable to Quine as statements of objective matters of fact. (It is indeed *plausible* to assume that the degree of symmetry that must exist even in the Quine-acceptable facts will ensure that there will be no significant conflicts between empirically adequate translation manuals; but plausibility is hardly an admissible consideration here.)

Quineans might be willing to concede that the symmetry would make it permissible – compatible with the Quine-acceptable facts – to say, of two rival translations provided by empirically adequate translation manuals, that they meant the same. But they would insist that it might nevertheless also be permissible to say that the two sentences did *not* mean the same, which is all the indeterminacy thesis needs. Nor is there anything in the argument from structure to show that the facts would inevitably prevent them from saying this.

So it seems that the argument from structure fails, and brings down with it the assumptions behind Davidson's suggestions. On the present interpretation Davidson would be merely begging the question against Quine, while on the earlier interpretation he would be misrepresenting Quine's thesis by gratuitously trivializing it. Either way, the thesis is unscathed.

University of Nottingham,
Nottingham NG7 2RD

© ROBERT KIRK 1985

REFERENCES

C. Boorse, 'The Origins of the Indeterminacy Thesis', *Journal of Philosophy*, 72 (1975), 369-87.

D. Davidson, 'Belief and the Basis of Meaning', *Synthese*, 27 (1974a), 309–23.
— 'Radical Interpretation', *Dialectica*, 27 (1974b), 314–28.
— 'Reality without Reference', *Dialectica*, 31 (1977), 247–58. Reprinted in Mark Platts (ed.), *Reference, Truth and Reality*, London etc.: Routledge and Kegan Paul (1980), to which page reference applies.
M. Dummett, 'The Significance of Quine's Indeterminacy Thesis', *Synthese*, 27 (1974), 351–97.
D. Lewis, 'Radical Interpretation', *Synthese*, 27 (1974), 331–44.
W. V. O. Quine, *Word and Object*, Cambridge, Mass., New York and London: M.I.T. and Wiley, 1960.
— 'Facts of the Matter', in *American Philosophy from Edwards to Quine*, R. W. Shahan and K. R. Merrill (eds.), Norman: University of Oklahoma Press, 1977, pp. 176–96.

SPATIAL ANALOGUES OF 'ANNIHILATION AND RE-CREATION'

By MICHAEL B. BURKE

IN *Metaphysics* (3rd edition, Prentice-Hall, 1983), as in his much discussed 1955 article on analogies between space and time ('Spatial and Temporal Analogies and the Concept of Identity', reprinted in *Problems of Space and Time*, ed. J. J. C. Smart, Macmillan, 1964), Richard Taylor gives as the spatial analogue of 'annihilation and re-creation' something that will here be called 'multiple location'. I will show that annihilation and re-creation comes in two varieties, that multiple location is the analogue of one (much the more problematic one), and that what will be called 'discontinuous extension' is the analogue of the other. (Except when modified by 'temporal', the terms 'location', 'extension', and 'place' will be short for 'spatial location', 'spatial extension', and 'spatial place'.) The object is not merely to expose the deficiency in Taylor's account, but to explore further the intriguing analogies between space and time.

Taylor's overall thesis is that the only point of disanalogy between space and time is that space has three dimensions while time has just one. In support of this thesis, Taylor considers several apparent disanalogies (such as that things can stop in space but not in time) and argues that they are apparent only.

One of these apparent disanalogies is that an object can be at the same place at different times, but, seemingly, cannot be at the same time at different places. Taylor's response is that an object *can* occupy different places at the same time. If, for instance, a billiard ball extends from one point in space to a second, it occupies the two places simultaneously. One is tempted to object that it is only a *part* of the ball that is in the first place and a different part that is in the second. But, replies Taylor, when an object occupies the same place at different times, it is only a *temporal* part of the object that is there at the first time and a different temporal part that is there at the second. The analogy is complete.

Thus far Taylor's reasoning is unproblematic, except, of course, for his reliance on the controversial doctrine of temporal parts.[1] The oversight occurs when Taylor undertakes to provide a spatial analogue for cases in which an object occupies the same place at different times without occupying *any* place at intervening times. Taylor writes:

> There was, we noted, one other way an object might conceivably be in the same place at two times, and that would be for it to be annihilated and then re-created at the same place. Some persons might want to regard this as the same object at both times, rather than two very similar objects, despite the lack of temporal continuity in it. But here the analogy is exceedingly simple. We need only to describe an object that exists simultaneously at two places and nowhere between, such as a billiard ball that is at once identically at both sides of the table ... [One might object that] we must say that these are two different billiard balls, however similar to each other they might be, simply on the ground that there is no spatial continuity between them ... But in the former case, too, in which an object is supposedly annihilated and then re-created, there is no temporal continuity ... One may suit himself whether there is an identity or a diversity of things in either case, but it would be arbitrary indeed to insist that there is an identity in the one case but a diversity in the other, the two situations being exactly analogous. (*Metaphysics*, pp. 67–8.)

As Taylor notes, it is questionable that annihilation and re-creation is even a logical possibility. I have argued elsewhere ('Cohabilitation, Stuff, and Intermittent Existence', *Mind*, 89, July 1980, pp. 391–405) that there actually *are* cases of what I call 'intermittent existence', but that is not the issue here. My aims are (a) to distinguish two types of annihilation and re-creation and (b) to identify the spatial analogue of each.

Taylor provides no explicit definition of 'annihilation and re-creation', evidently taking the expression to be self-explanatory. We will understand it as follows. An individual (object or event) is *annihilated and re-created* just in case there is a time t such that the individual does not exist at t, but does exist both earlier and later than t. (The *definiendum* properly applies only when the ceasing and recommencing to exist are *caused*, and perhaps only when even the *parts* of the individual cease to exist, but we needn't bother with these complications.)

Now if we employ the notion of a temporal part, we can distinguish two types of annihilation and re-creation (without claiming that either is even a logical possibility), which we will call 'intermittent existence' and 'recurrence'.

---

[1] This is the doctrine that physical objects (not just their histories), like events, have temporal parts – either because physical objects *are* events or because they are sums of momentary objects. On this doctrine, the '1985-part' of the White House is a temporal part of the individual of which the Oval Office is a spatial part. We will later look more closely at the notion of temporal parts.

The only one readily thought of is intermittent existence. An individual *exists intermittently*, we will say, just in case there is a moment or (more likely) a temporal interval which (1) contains no temporal part of the individual, (2) is bounded on the early side by a temporal part of the individual, and (3) is bounded on the late side by a later[2] temporal part of the individual. Imagine that a watch is completely disassembled for repair and that the same parts are later reassembled into their original order. Some would count this a case of intermittent existence. (Some may prefer a case involving toy building blocks, where it may be clearer that the disassembled object ceases to exist.)

What Taylor gives as the analogue of his undefined 'annihilation and re-creation' is what we will call 'multiple location': the occupation, by the whole of an individual (i.e., by all of the individual's spatial parts), of different places at the same time. Taylor's billiard ball is wholly at one end of the table and, simultaneously, wholly at the other. Now this may seem an obvious contradiction in terms. For is there not manifest verbal inconsistency in saying that all or even part of the ball is at the north end while at the same time saying that the *whole* of it is at the south end? Before proceeding, it will be useful to see that there is not. Afterwards we will ask whether multiple location is the spatial analogue of what we are calling 'intermittent existence'.

Consider the sentences below.

1 The whole of the ball is at the south end.
2 No part of the ball is somewhere other than the south end.
3 No part of the ball fails to be at the south end.

On a first reading, one might notice no difference in meaning between 2 and 3. It might appear that either would be a fully acceptable paraphrase of 1. It is this that explains why it might seem that multiple location is a non-starter. For the ball's being at the north end is undeniably incompatible, as a trivial consequence of the meanings of words, with the truth of 2. What must be noticed is that the ball's being at the north end is *not* thus incompatible with the truth of 3. All that would be required is that the ball be at the north end but that there be no part of the ball at the north end which is not *also* at the south end. If this latter is impossible, the reasons are metaphysical. To give Taylor's multiply located billiard ball a chance, we need only equate 1 with 3.

But now is multiple location, whether possible or not, the spatial analogue of intermittent existence? The answer is 'no'. In the case of the multiply located billiard ball, there is a spatial gap (containing no spatial parts of the billiard ball) between the two places occupied by the ball. So far there is analogy with the case of intermittent existence, where there is a temporal gap (containing no temporal parts of the intermittently existing individual) between earlier and later periods in which the individual exists. The point of *disanalogy* is this. Each spatial part of the doubly located billiard ball exists on both sides of the spatial gap (just as the ball itself does), but some of the temporal parts of an intermittently existing object exist on only one side of the temporal gap. (I say 'some' rather than 'all' because some of the *extended* temporal parts will straddle the gap. If, for example, an object exists throughout the 1980's, with the exception of June 1985, then the '1985-part' of the object is found on both sides of the temporal gap.) Succinctly, the disanalogy between multiple location and intermittent existence is this: every spatial part of a multiply located individual is itself multiply located; but not every temporal part of an intermittently existing individual itself exists intermittently.

---

[2] As will become clear, the word 'later' is needed here to exclude cases of what will be called 'recurrence'.

The spatial analogue of intermittent existence is what I will call 'discontinuous extension'. An individual is *discontinuously extended* just in case it exists while having two spatial parts which are connected by no line segment (straight or curved) whose every point is occupied by some spatial part of the individual. (Notice that a multiply located individual need not be discontinuously extended.) More simply (though less precisely), an individual is discontinuously extended if and only if its spatial parts are separated and yet *it* (not just its parts) exists. An example is a pipe whose stem and bowl are separated for cleaning, providing the pipe continues to exist. Another is a billiard ball half of which is at one end of a billiard table and half of which is at the other, if (which is doubtful) the ball still exists despite having been cut in half.

As in the case of intermittent existence, there is no need for us to take a stand on the possibility of discontinuous extension. The point here of interest is that discontinuous extension, whether possible or not, is the spatial analogue of intermittent existence. Indeed, it would be apt to call intermittent existence 'discontinuous *temporal* extension'. In the case of an intermittently existing object, there is a temporal gap (containing no temporal part of the object) between different periods in which the object exists, and no temporal part of the object exists (in its entirety) on both sides of the gap. In the case of a discontinuously extended object, there is a spatial gap (containing no spatial part of the object) between different places in which the object exists, and no spatial part of the object exists (in its entirety) on both sides of the gap. In the first case, there is a discontinuity in the temporal extension of the object; in the second, in its spatial extension.

We come now to the second type of annihilation and re-creation, which we will call 'recurrence', and which happens to be the temporal analogue of multiple location. Indeed, recurrence could aptly be called 'multiple *temporal* location'. An individual is multiply located (spatially) just in case all of its spatial parts exist in one region of space and, at the same time, also in another. We will say that an individual (object or event) *recurs* just in case all of its temporal parts exist in one region of time and, at the same place, also in another. Just as each spatial part of a multiply located individual is itself multiply located, so each temporal part of a recurrent individual is itself recurrent.

Like intermittent existence, recurrence satisfies our definition of 'annihilation and re-creation'.[3] In both cases, there is a time *t* such that the individual does not exist at *t*, but does exist both earlier and later than *t*. But only in the case of recurrence does each of the individual's temporal *parts* exist both earlier and later than *t*.

Most likely, the reader's first thought is that the notion of recurrence is straightforwardly incoherent. Is it not a plain contradiction in terms to suggest, say, that the 1985-part of a certain event might occur again in 2085? The answer depends on how 'temporal part' is defined. On the definition offered recently by Judith Jarvis Thomson ('Parthood and Identity Across Time', *The Journal of Philosophy*, 80, April 1983, pp. 206–10), the answer would be 'yes'. But this can be taken simply as an infelicity in Thomson's definition.[4] Consider Roderick Chisholm's similar, though less technical explanation of the temporal parts doctrine (which Chisholm, like Thomson, rejects).

> Just as an object that is extended through space at a given time has, for each portion of space that it occupies, a *spatial part* that is unique to that portion of space at that time, so, too, any object that persists through a period of time has, for each subperiod of time during which it exists, a *temporal part* that is unique to that subperiod of time. ('Problems of Identity', in *Identity and Individuation*, ed. M. Munitz, New York University Press, 1971, p. 15.)

Presumably, the clauses 'that is unique to that portion of space' and 'that is unique to that subperiod of time' are intended to mean 'that occupies all and only that portion of space' and 'that occupies all and only that subperiod of time'.

It is not clear whether the quoted passage is meant to provide an implicit definition of 'temporal part'. If it is, then Chisholm, too, has made the recurrence of temporal parts a non-starter. But notice that he has then done the same for the multiple location of spatial parts.

There is a simple fix. We need only rewrite the passage as follows.

> Just as an object that is extended through space at a given time has, for each portion of space that it (fully) occupies, a *spatial* part which is such that (1) it fully occupies that portion of space, and (2) all of it exists in

---

[3] To be perfectly accurate, the second type of annihilation and re-creation is instanced not by just any cases of recurrence, but only by those in which the occurrences are temporally separated.

[4] Thomson could accommodate recurrence simply by letting her 'T' range over *sets* of times.

that portion of space, so, too, any object that persists throughout a period of time has, for each subperiod of time throughout which it exists, a *temporal part* which is such that (1) it fully occupies that subperiod of time, and (2) all of it exists in that subperiod of time.

This formulation differs from Chisholm's only with respect to the two second clauses. It would be equivalent to Chisholm's, if those clauses read 'none of it exists ouside that portion of space' and 'none of it exists outside that subperiod of time'. But our formulation has the merit of leaving it an open question whether recurrence and multiple location are possible. Our definition of '1985-part' would contain not the clause 'none of which exists outside 1985', but rather the clause 'all of which exists in 1985' (meaning 'none of which fails to exist in 1985'). The latter, unlike the former, leaves it to metaphysics whether a 1985-part might exist again in 2085.

As with multiple location, it is hard to conceive anything that might plausibly be counted a case of recurrence. Nietzsche's 'eternal recurrence' is a candidate. And one might hold successive blasts of a whistle to be the same one blast (token) occurring at different times, just as one might, with equal implausibility, hold balls on opposite sides of a billiard table to be the same one ball existing at different places. If the whistle blasted from 1:00 to 1:05 and again from 2:00 to 2:05, one might say that the 1:00-part of the blast recurred at 2:00, that the 1:01-part of the blast recurred at 2:01, and so on.

In conclusion, there are two types of annihilation and re-creation: intermittent existence and recurrence; their spatial analogues are, respectively, discontinuous extension and multiple location; although all four are of questionable possibility, none, properly defined, is impossible as a trivial consequence of its definition.

Indiana University,
425 Agnes Street,
Indianapolis, Indiana 46202,
U.S.A.

© MICHAEL B. BURKE 1984

THE MARTYR'S DILEMMA

By F. C. T. MOORE

IT is sometimes stated or implied that the Prisoner's Dilemma arises only for rationally self-interested behaviour. Hillel Steiner, for example, has said ('Prisoner's Dilemma as an Insoluble Problem', *Mind* XCI (1982), pp. 285-6) that the dilemma depends upon the statement that 'all persons pursue rationally self-interested strategies'. Commentators on Steiner's article appear to have accepted this widespread view (see Gordon, D., 'Is the Prisoner's Dilemma an Insoluble Problem?', and Porter, J. P., 'Relevant Interest and the Prisoner's Dilemma', *Mind* XCIII (1984), pp. 98–102). Thus the question of the rationality of altruism may be held to be settled – or begged – by the dilemma. Indeed, in the usual matrix (for two players A and B and two choices C and D) where $T > R > P > S$ (the latter, of course, being a preference ordering), C is commonly described as the 'altruistic' (but irrational) choice.

		B	
		C	D
A	C	R,R	S,T
	D	T,S	P,P

Figure 1

In fact, moralizing readings of the Prisoner's Dilemma can be quite elaborate: thus C is read 'cooperate', and R is the 'reward' obtained by A and B for their cooperation; D is read as 'deviation' or 'defection', and P is the 'punishment' both receive for their wickedness; in the case of asymmetrical choices, the player gaining T is described as a 'traitor', while the loser (who takes S) is a 'saint' or even a 'sucker'. David Lewis has claimed it to be 'rat'ional to 'rat' (in 'Prisoner's Dilemma is a Newcomb problem', *Philosophy and Public Affairs* 83 (1979), pp. 235–40), while Anatol Rapoport in an early article ('Formal Games as Probing Tools for Investigating Behaviour Motivated by Trust and Suspicion', *Journal of Conflict Resolution* VII (1963), pp. 570–9) described a player's repeated cooperation in iterated games (in spite of defection by the other player) as a 'martyr run'.

Various authors have questioned such readings of the Prisoner's Dilemma, including recently M. C. Geach (ANALYSIS 44.1, January 1984, pp. 46–8: 'a similar sort of case can arise even when people agree about the desired effect, and are not selfish'). Similarly, J. W. N. Watkins has maintained, in 'Self Interest and Morality' (Körner, S. (ed.), *Practical Reason*, Oxford, Blackwell, 1974, pp. 67–77), that the dilemma arises for moral agents who give equal weight to their own and the other's preferences. And Derek Parfit in 'Is Commonsense Morality Self-defeating?' (*Journal of Philosophy* 76.10, October 1979, pp. 533–45), applies Prisoner's Dilemma reasoning to cases where the parties are pursuing different goals which each flow from the same moral theory, in order to construct an argument that any such theory is self-defeating. My claim is more general, that Prisoner's Dilemma arises for any actions whatsoever in whose case the outcome depends upon the interplay of different actions – it is thus a problem for the theory of social action, and does not in any way suppose or show the irrationality of altruism.

Note also recent work on applications of Game Theory to the evolution of behaviour in Maynard-Smith, J., *Evolution and the Theory of Games*, CUP, 1982, and 'Game Theory and the Evolution of Behaviour', *The Behavioural and Brain Sciences* 7 (1984), pp. 95-125, including pp. 114-15, the commentary by Rapoport, 'Game Theory without Rationality'. These applications assume the possibility of different strategies, or 'behavioural phenotypes', being present in a population (whether because some individuals are genetically determined to do x, and others to do y, or because an individual is genetically determined to do x and y with certain probabilities, or because an individual tends to do x and y with certain probabilities as a result of learning). It is then possible to calculate whether there is an 'evolutionarily stable strategy' (pure or mixed) by interpreting payoff in terms of Darwinian fitness, i.e. in terms of the numbers of offspring produced by organisms with various strategies. Efforts have been made using this approach to offer an empirical explanation of the phenomenon of altruism. But it is more relevant to the present argument that the approach itself theoretically allows any behavioural tendencies to occur. We might, for example, have entire populations of doves – or martyrs. The empirical questions are: which behavioural phenotypes actually occur, and which have survival value.

It is not my aim here to say anything about solutions to the Prisoner's Dilemma, but rather to establish a point about its character. Indeed, one interesting line of solution was developed formally by Howard, N., in 'The Theory of Meta-Games' and 'The Mathematics of Meta-Games' in *General Systems* 11, 1966, pp. 167-200, and explained and elaborated on by Pörn, Ingmar, *Action Theory and Social Science: some formal models*, Dordrecht, Reidel, 1977, pp. 98-102. This line of solution involves the introduction of policies in the first-level meta-game (by which *b* makes his choice of C or D conditional on *a*'s choice of C or D), and then in the second-level meta-game *a* can make his choice of C or D conditional upon *b*'s choice of *policy*. If this line of reasoning is accepted, it turns out that a cooperative policy is rationally preferable to a non-cooperative policy (see also Rapoport, A., 'Escape from Paradox', *Scientific American* 217 (1967), pp. 50-6).

My aim here is rather to support the view that the Prisoner's Dilemma is morally neutral by arguing that it arises out of interdependent choices under uncertainty, whatever the moral character of the objectives of the agents, and – in particular – independently of whether those objectives are self-interested or altruistic in character. The rational choice may commonly have been seen as a 'selfish' or self-interested one, but this is only because self-interested objectives happen to have been selected in illustrating the dilemma.

It is sufficient proof of this contention to construct another instantiation of the same problem in which the objective is admittedly altruistic.

Three people, A, B and C, are in a dug-out. A hand-grenade lands close to and between A and B, but some distance from C. This is how the situation appears to A and B: if A and B both throw themselves away from the grenade, all three will be injured. If A throws himself on the grenade, he will die, and B will be injured, but C will be unscathed. Correspondingly, if B throws himself on the grenade. If both A and B try to throw themselves on the grenade, both will be seriously injured, but C will be unscathed.

If A is a martyr, all he will be concerned with is the fate of the others: that is, in the case of the four possible *outcomes* as they affect the three men in the dugout, we measure the *payoffs* from A's point of view by ignoring the *outcome* for himself. Now let us suppose that in fact both A and B are martyrs. Let them accordingly share the following preference ordering over the possible outcomes (for both of them, as explained, the outcome for only two people out of the three in the dugout is relevant):

4. worst: 1 dead
3. very bad: 1 seriously injured
2. bad: 2 injured
1. least bad: 1 injured

What then should A do? Consider this diagram:

		B	
		saves himself	sacrifices himself
A	saves himself	2i/2i	1d/1i
	sacrifices himself	1i/1d	1si/1si

Figure 2. Outcomes as judged by A and B, each ignoring himself
(i = injured, d = dead, u = unscathed, si = seriously injured)

Given A's preferences specified above (i.e. 1 least bad, 2 bad, 3 very bad, 4 worst), he faces the following situation. If he sacrifices himself, there are two possible outcomes – 3 very bad (if B does the same), or 1 least bad (if B tries to save his skin); but if he saves his skin, the two possible outcomes are – 4 worst (in case B sacrifices himself) or 2 bad (in case B also saves his skin). By minimax, as a martyr, A should sacrifice himself, since the outcome cannot be worse than 3 very bad, and there is a chance of 1, the least bad outcome. He should not save his skin, since though there is a chance of 2 the (merely) bad outcome – which is collectively the best for two martyrs, there is also a chance of 4, the worst outcome, if the other sacrifices himself. Let me emphasize again that if A is a martyr (that is, if his preference ordering of the outcomes is the one already described) then it would be rational for A, using minimax, to sacrifice himself; but if both A and B are martyrs the best result for them collectively, i.e. the outcome which would produce the best net result taking both their preferences into account, would be achieved by both saving their skin, for there will then be two injured, which both agree to be better than one seriously injured, or one dead.

Thus if both A and B are martyrs, they will in 'rationality' both sacrifice themselves, producing an outcome 3, which is very bad (i.e. one seriously injured, from either point of view); but irrational martyrs, both of whom saved their skin, would actually produce the best net result. This can be checked on the following diagram:

		B	
		saves himself	sacrifices himself
A	saves himself	A B C i i i	A B C i d u
	sacrifices himself	A B C d i u	A B C si si u

Figure 3. Actual outcomes for all three

I conclude that we can readily construct a dilemma formally identical to the Prisoner's Dilemma, but in which the parties have altruistic objectives, in the sense of ignoring their own interests. My dilemma could be called the 'Martyr's Dilemma' – except that if my argument here is cogent, the two dilemmas are in fact just the same dilemma in different instantiations. I am uncertain whether the example tells us anything about martyrdom; but it does, I believe, prove that moralizing interpretations of the Prisoner's Dilemma are wrong.

University of Hong Kong,
Pokfulam Road, Hong Kong

© F. C. T. Moore 1985

WHAT IS A FUNDAMENTAL ETHICAL DISAGREEMENT?

By WAYNE WASSERMAN

PHILOSOPHERS often distinguish between philosophical and descriptive ethical relativisms. The latter is held to be within the province of science, and 'first among its theses, without which the others would lose interest, is that the values, or ethical principles, of individuals conflict in a fundamental way'.[1]

But, as I shall argue, it is hard to say what such theses mean, for it is unclear what counts as a fundamental ethical agreement or disagreement. And reflecting on the problems will not only show the inadequacy of some standard accounts (e.g., Brandt's, Paul Taylor's) of the notions, but will show implausible attempts to draw a sharp line between descriptive relativism and ethical theory.

Richard Brandt's classical characterization will serve as a useful target:

> to say that a disagreement is 'fundamental' means that it would not be removed even if there were perfect agreement about the properties of the thing being evaluated ... Thus it is not necessarily a case of fundamental disagreement of values if one group approves of children's executing their parents at a certain age or state of feebleness whereas another group disapproves of this very strongly. It may be that in the first group the act is thought necessary for the welfare of the parent in the afterlife, whereas in another group it is thought not to be. This disagreement might well be removed by agreement about the facts ... both parties might subscribe, now, to the principle 'It is right for a child to treat a parent in whatever way is required for the parent's long-run welfare.' The disagreement might be simply about the implications of this common principle, in the light of differing conceptions of the facts. (Brandt, op. cit., p. 75.)

The account requires elaboration. For one thing, Brandt is presumably interested in counterfactual conditions of agreement about relevant facts, and this notion needs work. In another context Brandt has attempted to explain the notion of facts being relevant to moral judgment (*A Theory of the Good and the Right*, Oxford 1979, pp. 12, 111-12). Perhaps his account there could be useful in the present context. But for the present I think it enough to say that the problem of getting an adequate account of this may be difficult. Indeed, my fear is that agents with different ethics may have profoundly different conceptions of what counts as a fact relevant to moral assessment, and thus that any account of relevance of facts will be relative to an ethical perspective.

A second matter is that the account needs to be clarified to rule out uncertainty about the following. Does Brandt require that a non-fundamental disagreement be one which would disappear under any and all possible ways of agreeing factually about the situation in question? Or is the requirement weaker? For example, is it enough that the disagreement would end if the first and second agents were to agree either on the first or second's factual beliefs? On this view a non-fundamental ethical disagreement could be one which would dissolve were the parties to agree on the first's way of viewing the facts, even though it would not dissolve were they to agree on the second's factual assessment. A different approach would be to require that both of these ways of coming to agree factually would end the ethical disagreement. There are probably other possibilities, but let us focus on these.

---

[1] Richard Brandt, 'Ethical Relativism' in *The Encyclopedia of Philosophy*, ed. Paul Edwards (Macmillan, 1967), Vol. III, p. 75. Brandt offers much the same account in *Ethical Theory* (Prentice-Hall, 1959).

It is not clear which way Brandt would have wished to go. But Brandt's choice of examples and his interest in the anthropology of ethics suggest that his discussion here is informed by a particular controversy. In response to studies showing the great differences between different societies' ethics, some philosophers and social scientists attempted to undercut the force of sceptical claims based on those discoveries by urging that there were ethical universals. Such shared basic values or ethical principles were said to underlie the surface disagreements between cultures; the disagreements were supposed to be about derivative ethical beliefs, and were said to arise only because the shared basic valuings or beliefs were combined, by different cultures, with different factual beliefs. Were this the context of Brandt's writing, one would expect that he would have opted for something like the first view described above. For if two agents' disagreement was supposedly non-fundamental, and if a non-fundamental disagreement was one where the agents' basic values were the same, then any way of agreeing on the facts should lead the agents to the same derivative judgments.

The last comments are conjecture. Fortunately, nothing in the balance of my discussion turns on the differences between the possible ways of understanding Brandt. This is because any attempt to make the distinction between fundamental and non-fundamental ethical disagreement solely in terms ethical agreement under conditions of contrary-to-fact agreements about factual matters will be inadequate. I turn now to indicate some of the problems.

Suppose X and Y shared a single basic ethical principle and had exactly the same factual beliefs. Suppose that X but not Y made a simple error of deductive reasoning when combining ethical and factual beliefs. The two would then be in a condition of factual agreement, yet they would still disagree ethically. So on Brandt's account this would have to be characterized as a fundamental disagreement, though it surely is not one.

This problem cannot be avoided simply by dismissing all disagreements arising from differences of reasoning. Plainly, moralities can as fundamentally differ in principles of reasoning as in points of principle. For example, X and Y might begin with the same ethical and factual beliefs, but X and not Y might characteristically employ a form of analogical ethical reasoning akin to that characteristic of legal reasoning. Here the disagreement resulting from differences of reasoning seems fundamental.

Such differences in preferred patterns of ethical reasoning may, but need not, be accompanied by differences in metaethical beliefs about correctness of reasoning in ethics. This suggests that beliefs which are neither clearly factual nor ethical may produce intuitively fundamental ethical disagreements. This is clear in the case of metaethical beliefs about correctness of ethical reasoning. If the agents discussed in the preceding paragraph initially had the same factual and ethical beliefs, and had similar characteristic patterns of ethical reasoning, the adoption by X and rejection by Y of the belief that it is appropriate to employ the kind of reasoning described above would eventually lead to much the same situation as just described. How would Brandt have us treat such beliefs for purposes of using his account? Are they to count as ethical or factual?

But presumably not all beliefs about correctness of reasoning should be treated similarly when trying to develop an account like Brandt's. There is a considerable difference between an ethical dispute arising from a metaethical difference about correctness of peculiarly ethical reasoning and one arising from a difference about what is and is not a correct way of using sophisticated statistical arguments.

Besides beliefs about correctness of reasoning there may be other normative beliefs which are relevant to the discussion, but which do not fit easily into Brandt's ethical *versus* factual dichotomy. For example, two agents who seem to agree ethically may come to adopt conflicting positions of public policy because, although they agree that scarce public resources should be allocated to protecting great works of art, they disagree violently about what is a great work of art. Here I find myself tempted to say that the dispute is fundamental, although it seems to arise from a difference of aesthetic taste.

The preceding observations suggest that the initial distinction between factual and ethical claims needs to be augmented if an analysis like Brandt's is to work. As I do not see how to make the necessary emendations, I am in doubt as to the adequacy of the sort of approach Brandt has adopted.

In addition the analysis simply speaks of the cessation of ethical disagreement under a counterfactual condition of factual agreement. More satisfactory approaches would make finer distinctions among different situations of disagreement and different conditions of counterfactual agreement, and they would take into consideration the need to tailor different clauses to different ethics.

Consider first some situations of ethical conflict:

d1 X and Y disagree about ethical claim e and non-ethical claim F, and they agree that coming to agree on F would end their disagreement about e.

d2 X and Y disagree about e and F, and agree that ceasing to disagree about F would not end their dispute about e.
d3 X and Y disagreement about e and F, but neither has an opinion about what would happen were they to agree on F.

There are other situations of conflict, but let us focus on these. Consider next some conditions of contrary-to-fact ethical agreement:

c1 Were X and Y now to cease differing about F, they would now cease differing about e.
c2 While now ceasing to disagree about F would not end their disagreement about e, had they never disagreed about F they would not now disagree about e.

Suppose now that we first consider the fundamentality of disagreement from the perspective of a perceptual intuitionist, one who holds that a quasi-visual sensing reveals the presence of non-natural ethical properties analogous to yellowness. From such a perspective nearly all disagreements are fundamental, for virtually none can be explained away as a disagreement arising from a dispute about the facts.

On the other hand, consider a highly Cartesian ethic which holds moral belief to be an entirely conscious, rational grasping of a few general ethical principles, and the combining of these with factual beliefs to deduce derivative ethical beliefs. All these mental acts and contents would be transparent to consciousness, so that reflection by an agent would reveal to her what was ethically basic and what was derived. On such a conception the fundamentality or non-fundamentality of disagreement would be open to the agents' introspection. Any non-fundamental disagreement would be resolvable by elimination of factual disagreement. I conclude that from the perspective of such an ethic only a d1–c1 disagreement would be non-fundamental.

A naturalistic ethic which stressed the role of feelings and instincts and rejected the Cartesian claims about the rationality and transparency of ethical belief would be very different. A non-fundamental disagreement need not be a d1 disagreement, for the agents involved might not know their own minds, as it were. Furthermore, a non-fundamental disagreement might be open to a c2 resolution, unlike the case of the Cartesian ethic. The following line is, after all, plausible given the view in question. Once the basic and fundamentally similar natural tendencies of two agents are subjected to the differing influences exerted by different factual beliefs, the agents may not easily return to ethical agreement, even if they overcome the factual disagreement. While perhaps morally irrelevant, there are strong forces which make it difficult to admit one has erred ethically. These tendencies would sustain a non-fundamental disagreement.

It seems, then, that judgments of fundamentality of agreement and disagreement are relative to conceptions of ethics. Since many of us at least on occasion are uncertain of which conception of ethics to accept, and since we sometimes are concerned with how things stand from the perspectives of others, perhaps it would be more useful to search for analyses of the form 'relative to conception of ethics C, X and Y fundamentally differ iff ___', rather than to seek a single theory-independent analysis.

Let us turn to a different, serious problem. Intuitively the counterfactual clause was included to screen out disagreements due to non-ethical differences. Yet it may also screen out cases which are intuitively fundamental disagreements. Interestingly, the counterexample I offer also works against an account of fundamental ethical disagreement offered by Paul Taylor, for Taylor in effect suggested that agreement on a single necessary and sufficient condition was sufficient to yield fundamental ethical agreement (Paul Taylor, *The Principles of Ethics*, Dickenson 1975, pp. 16–18).

Suppose X and Y fervently believe that something is wrong if and only if The Deity does not approve of it. Suppose also that they 'learn' about The Deity's preferences through 'dream visitations' which are only projections of their own deepest, hidden values. Having thus discovered that The Deity approves of abortion, X employs 'the divine biconditional' and discovers that abortion is permissible, while Y similarly discovers that abortion is wrong.

Now the connections between the values and factual beliefs of such an agent might well be so intimate that the only way of producing a change of belief about The Deity's preferences would involve changing the agent's values. After all, factual beliefs are sometimes determined by and are mere reflections of strongly held values.[2] X and Y would then cease disagreeing about The Deity if and only if they were to undergo basic changes in their values. So the following would be true: were X and Y to agree on the factual claims about The Deity they would not disagree about the morality of abortion. Thus on Brandt's account of fundamentality their difference would count as non-fundamental.

But this is wrong, and no more convincing than to say that the dispute is non-fundamental because it would end if they were to cease disagreeing ethically.

This serious problem suggests that a very different approach may be in order. Rather than including a counterfactual condition which builds into the analysis conditions for evidencing judgments of fundamentality, it might pay to move to an analysis which straightforwardly rules out disagreements arising from non-ethical differences, e.g. to something like the following:

C X and Y's disagreement about ethical claim e is non-fundamental iff the relevant cause of the disagreement is a non-ethical difference between the two.

---

[2] For graphic illustrations of the determination of factual belief by ethical belief see Melvin J. Lerner's *The Belief in a Just World* (New York: Plenum Press, 1980).

This approach is probably problematic on a number of accounts, but because it forces us to think in terms of the mechanisms producing ethical beliefs it has the virtue of forcing us to confront a problem which had dogged us all along.

Our pursuit of fundamentality of agreement or disagreement is the result of a desire to go beyond the mere listing of disagreements. We want to know if the relatively observable facts suggest some more profound thesis about ethical disagreement, e.g. one relevant to the adequacy of sceptical theses. Since something like this is at issue, we want an account of fundamentality which is not only conceptually adequate, but is of use in making decisions about whether or not a 'deeper' thesis is correct. And this brings us to the problem.

Even given an adequate analysis, decisions have to be made about whether or not the defining conditions obtain. And it is hard to see how fully adequate judgments of that sort can be obtained in the absence of fairly definite views about the mechanisms producing ethical beliefs. After all, we probably will need to make judgments about counterfactual situations of the sort described by Brandt, and it seems likely that judgments about such situations will need to be based on causal generalizations. It is unlikely that fully adequate generalizations of that sort can be obtained in the absence of fairly definite views about the mechanisms of ethical belief. But forming definite views about those mechanisms involves taking definite positions in ethical theory.

Any interesting claim of descriptive ethics must involve claims about the fundamentality of ethical agreement. But we have seen the inadequacy of the classical characterization of that notion. We have also seen reason to think that real progress will require doing philosophical ethics. Analyses of what counts as a fundamental disagreement need to be tailored to different conceptions of ethical belief, and once we have obtained such analyses they will be fully usable only given background beliefs about the nature of ethical belief.[3]

University of Nebraska,
Lincoln, Nebraska 68588, and
Hollins College,
Virginia 24020, U.S.A.

© WAYNE WASSERMAN 1985

---

[3] I thank Mark Timmons, the Editor, Tracy Bacigalupo, and especially Phil Hugly.

IS MORAL RELATIVISM CONSISTENT?

By PHILIP HUGLY *and* CHARLES SAYWARD

IN Tehran it is prohibited to proclaim publically the non-existence of Allah. In Lincoln, Nebraska (U.S.A.) this is not prohibited. Some have suggested

(X) One's moral judgment is valid just in case it agrees with the moral code of one's social group.

Others have suggested that (X) is inconsistent. They reason as follows: The judgment

(J) It is wrong to proclaim publically the nonexistence of Allah

made by a native of Tehran is valid (according to (X)). The same judgment made by a native of Lincoln, Nebraska, is not valid (according to (X)). This is inconsistent. So (X) is false.

There is an obvious reply to this line of reasoning. It is not obvious that no judgment can be both valid and not valid. Compare

No fence is both green and not green.

A fence can be green at one time and not green at another time. Similarly, a sentence might be true relative to person p and time t and not true relative to person q at time s. The sentence 'The door is open' is true (in English) relative to person p and time t just in case the door p has in mind at t is open at t.

Might not 'valid' as applied to moral judgments behave similarly? As a first shot in adapting (X) to this idea we have

(Y) (J) is valid relative to person p and time t just in case (J) agrees with the moral code of the social group to which p belongs at t.

Clearly this will not do because p might belong to more than one social group at time t.

We might try a simple emendation to get around this: (J) is valid relative to p at t just in case (J) agrees with the moral code of a social group p belongs to at t. But if (J) agrees with norms of social group G_1 and p belongs to G_1 at t and (J) does not agree with the moral code of G_2 and p belongs to G_2 at t, then (J) is valid relative to p at t and (J) is not valid relative to p at t. So this will not do, either.

There is a way of getting around this.

(Z) (J) is valid relative to person p at time t just in case (J) agrees with the moral code p has at t.

Could p have more than one moral code at a given time?

To answer this we need to be more explicit about what is to be counted as a moral code. A moral code shall be viewed here as a set of moral norms closed under the consequence relation. These norms can appear in propositional form or as imperatives, e.g. 'In circumstances C it ought to be the case that A' or 'Do A in circumstances C'. The set consisting of the ten commandments plus consequences is an example of a moral code. A moral judgment that A ought to be done agrees with a code at a time just in case the code contains a norm which says that A ought to be done in such and such circumstances and the circumstances are such and such at that time.

Now suppose C_1 and C_2 are distinct sets of norms which say what morally ought to be done in various circumstances. Suppose person p adheres to both at time t. Suppose there is no other set p also adheres to at t. Then $C_1 \cup C_2$ is *the* moral code p has at time t. So, to answer the question raised after (Z): a person can have at most one moral code at a given time.

Does every person have a moral code at any time? Just because a person cannot articulate the moral code he has at t we cannot infer that he has none. But there are people who at least *say* they do not have a moral code: 'I have no dealings with *oughts* and *oughtn'ts*. I forswear all such talk. I deal only with *needs* and *wants*.' We see little reason for not taking such claims at face value. The moral relativist ought to allow for the possibility that, for some person p and time t, p has no moral code at t.

According to (Z) both these things are true for such a person p at time t:

(1) (J) is not valid relative to p at t
(2) $\neg$(J) is not valid relative to p at t.

There is nothing inconsistent or incoherent about this. And it seems to be just the sort of result one would welcome if one were a moral relativist. If a person has no moral code then no moral judgment is valid for that person.

A similar result occurs if p has a moral code at t, but an incomplete one. If the code is incomplete there will be judgments such that neither they nor their negations agree with the code. Thus if the code p has at t is incomplete, and, in particular, incomplete with respect to (J), then both (1) and (2) will be true. Again, no problem.

But suppose p's code at t is inconsistent. Then all judgments (or at least all judgments formulable in a language in which the code is formulable) would agree with that code. Thus, for such a p and t, both

(3) (J) is valid relative to p at t

and

(4) $\neg$(J) is valid relative to p at t

are true. Is not this inconsistent?

Compare

(i) Statement S and its negation are true relative to some person p at some time t.

(ii) Statement S and its negation are both theses of some theory T.

(i) is inconsistent but (ii) is not. Of course the theory T mentioned in (ii) would be inconsistent. But that is another matter. It is not inconsistent to say that such a theory is inconsistent. The view of moral relativism we are considering says there is no moral analogue to truth. *Valid moral judgment* does not correspond to *true statement.* The closest analogue one can come up with is *thesis of a theory.*

The moral relativist might say this: Nothing is simply a thesis. A statement is always a thesis *of a theory.* Similarly, no moral judgment is just valid. It is valid relative to a code C. A moral judgment is valid relative to C just in case it agrees with C. A moral judgment is valid relative to a person at a time just in case the judgment is valid with respect to the code that person has at that time. If the person p holds an inconsistent code at t then of course both a judgment and its negation will be valid for p at t. But this in itself is not inconsistent. It is a result that simply records the fact that p's code is inconsistent.

The claim that an opponent of moral relativism should focus on at this juncture is this:

(W) There is no system-independent notion of moral correctness.

Now here is how an objection to (W) might proceed. Consider the notion of consistency of a moral code which is being appealed to in getting (3) and (4). It is this:

> Code C is consistent just in case for no norm N is it the case that N and its negation belong to C.

Now what is the negation of a norm? Is this just a syntactical notion? It does not seem so. For inconsistency of a code is not necessarily an objectionable feature if all that is meant by this is that something of one shape, say,

> In all circumstances it is wrong to proclaim publically the nonexistence of Allah

and something of another shape, say,

> In some circumstances it is not wrong to proclaim publically the nonexistence of Allah

belong to the code. What makes inconsistency objectionable is that not both of these norms can be true (or valid or correct or whatever you want to call it). So (the objection continues) a semantical notion of consistency of a moral code is possible only if truth or something analogous is applicable to the norms which make up a moral code. And this notion of truth is not system relative. It is the notion of true in a language.

The objection seems to us to be a serious one. But we do not think it is correct. For if it were, the same objection would show that a system-independent notion of correctness is applicable to all sorts of norms besides moral norms. For example, suppose

> She is beautiful

is asserted of some woman by an Irish person and denied of the same woman at the same time by an Italian person. It is plausible to suppose that the Irish person and the Italian person are appealing to different systems of standards and that 'She is beautiful' is correct relative to one system of standards and incorrect relative to another system.

Is there, in addition, a system-independent notion of correctness according to which one can say the one person made a correct judgment and the other an incorrect judgment? If the objection to (W) just propounded were sound we should have to say so. For aesthetic norms stand in logical relations. And systems of such norms can be consistent and inconsistent, complete, incomplete, and so forth.

The question before us is whether the fact that norms admit of logical properties and relations *shows* that a system-independent notion of correctness is applicable to them.

Our answer is that it does not show this. A system-relative concept of truth is all that is needed for explicating such logical properties as logical truth, consequence and inconsistency. This holds true even if we are talking about an ordinary first order language.

Let L be the first order language Mates describes in *Elementary Logic* (Oxford 1965), chapter 3. Call K a system if K is a set of atomic sentences of L. Truth relative to K is defined as follows:

1. If ϕ is an atomic sentence of L then ϕ is true relative to K iff $\phi \in K$.
2. A sentence $\neg\phi$ is true relative to K iff ϕ is not true relative to K; similarly for the other connectives.
3. A sentence $(\alpha)\phi$ is true relative to K iff $\phi\alpha/\beta$ is true relative to K, for all names β. ($\phi\alpha/\beta$ results from ϕ by replacing each free occurrence of α by an occurrence of the name β).

Now a sentence ϕ of L is a logical truth just in case ϕ is true in all complete interpretations of L. (A complete interpretation is one in which each domain element is named.) From this fact it can be shown that a sentence is logically true just in case it is true relative to all systems.[1] So a system-relative notion of truth suffices for explicating logical truth, and, thus, also consequence and inconsistency.

---

[1] The restriction to languages such as Mates's with denumerably many names is inessential. It can be shown by a proof of a length and complexity inappropriate to this paper that for each language L, with or without names, and each sentence ϕ of L, ϕ is true in an interpretation of L if and only if ϕ is true relative to some set of atomic formulas of L, and conversely.

Let C_1 and C_2 be distinct moral codes formulated in English. Let C_1 contain a norm N and C_2 its negation. The moral relativist is saying that if both codes are consistent, then, in the strongest sense of correctness applicable to moral norms, they are also both correct in the sense that they contain only correct moral norms. Consider the same thing said of physical theories: 'Let T_1 and T_2 be two consistent physical theories. Let T_1 contain a statement S and T_2 its negation. Then, in the strongest sense of correctness applicable to physical statements, both theories are correct in the sense of containing only correct physical statements.' If we believe that the physical statements of English are true in English we will reject the second claim. We will hold that the strongest sense of correctness applicable to physical statements is not system-relative.

The moral relativist denies that there is any corresponding sense of correctness applicable to moral norms. That is, there is no notion of moral correctness which is not system-dependent. This position may not be true, but there is no strictly logical basis for refuting it.

University of Nebraska-Lincoln,
1010 Oldfather Hall,
Lincoln, Nebraska 68588-0321

© PHILIP HUGLY *and* CHARLES SAYWARD 1985

'OUGHT TO HAVE' AND 'COULD HAVE'

By WALTER SINNOTT-ARMSTRONG

MANY arguments concerning practical reasoning assume that 'ought' implies 'can'. More specifically, most such arguments depend on the assumption that this implication is a semantic implication, such as an entailment or a semantic presupposition, rather than a pragmatic implication, such as Gricean conversational or conventional implicature [1]. If the implication is only pragmatic, it can be true that an agent ought to do something even if he or she cannot do it. Thus, any argument from the premise that an agent cannot do something to the conclusion that it is not true that an agent ought to do it is invalid unless the implication from 'ought' to 'can' is semantic.

There are, however, several reasons for denying that 'ought' semantically implies 'can'. I have given counterexamples elsewhere (see [2]). In this article, I will give a more technical argument for the same conclusion.

Semantic implications relate only sentences or propositions. I will speak of sentences, but my arguments do not depend on this. Thus, the claim that one operator semantically implies another operator must be short for the claim that sentences with one operator semantically imply sentences with the other operator. When tenses occur in sentences with operators, the tenses can apply either to the operator itself or to the sentence inside the scope of the operator.

If sentences with one operator semantically imply sentences with another operator, then the implication must hold even when the sentences within the scope of the operators are about the past. For example, 'knows' entails 'believes', so

(KD) John knows that Bill has a dog

entails

(BD) John believes that Bill has a dog,

so (prefixing 'P' to what is past)

(KPD) John knows that Bill had a dog

entails

(BDP) John believes that Bill had a dog.

However, (KPD) does not entail

(PBPD) John believed that Bill had a dog

or

(PBD) John believed that Bill now has (or would have) a dog.

Present knowledge of the past does not entail past belief. Thus, semantic implications between operators hold for sentences about the past without requiring the past tense to apply to the entailed or presupposed operator.

The relation between 'ought' and 'can' is not like this, because the past tense usually must apply to the operator 'can' when the past tense does not apply to 'ought' but to the sentence inside the scope of 'ought'. For example,

(X) I ought to have mowed your grass

can be interpreted either as

(OPM) I now ought to have mowed your grass in the past

or as

(POM) It was true in the past that I ought to mow your grass.

These interpretations differ in whether the past tense applies to the operator or to the sentence inside the scope of the operator. This difference makes these interpretations true in different circumstances. For example, suppose I promise on Monday to mow your grass by Saturday. (POM) is still true on Thursday and on all future days, because it was true on Tuesday that I ought to mow your grass by Saturday. In contrast, (OPM) is not true on Thursday, because it is not yet true that I ought to have mowed your grass in the past (before Thursday), since I promised only to mow it by Saturday. Since it does not seem true to say (X) on Thursday, and (POM) is true but (OPM) is not true on Thursday, (OPM) seems more plausible as an interpretation of (X) at least in this situation on Thursday.

If 'ought' semantically implied 'can', 'ought to have' would semantically imply 'can have', so (X) would semantically imply

(Y) I can have mowed your grass.

'Can have' seems odd, but that supports what I will say. (Y) can be interpreted either as

(CPM) I now can bring it about that I did mow your grass

or as

(PCM) It was true in the past that I could mow your grass.

These interpretations differ in whether the past tense applies to the operator or to the sentence inside the scope of the operator. (CPM) and (PCM) are not true in the same circumstances. For example, suppose on Saturday I have the ability to mow your grass on Saturday, but I do not mow your grass on Saturday. (PCM) is true on Sunday (and all future days), since on Saturday I had the ability to mow your grass on Saturday. In contrast, (CPM) is not true on Sunday, because it is too late to change the fact that I did not mow your grass on Saturday (or ever in the past).

These distinctions make it clear that (OPM) can be true in situations where (CPM) is false. For example, suppose I promise on Monday to mow your grass by Saturday, and you do not cancel my obligation, and I have no excuse or justification for not mowing your grass by Saturday, but I do not mow your grass by Saturday. (OPM) is true on the following Sunday, because I ought to have mowed your grass by Saturday and thus before Sunday. However, (CPM) is not true on Sunday, because I cannot on Sunday do anything that will bring it about that I did mow your grass by Saturday. Since (OPM) can be true when (CPM) is false, 'ought' does not semantically imply 'can'.

An opponent might respond that (OPM) might still semantically imply (PCM). However, this implication is not like the semantic implications discussed above. Furthermore, the claim that (OPM) semantically implies (PCM) is not enough for many of the arguments that assume that 'ought' implies 'can', because this claim allows the possibility that an agent ought to do something that the agent (now) cannot do.

A second possible response is that (OPM) implies (PCM) rather than (CPM) only because nobody can change the past, so no judgment like (CPM) can ever be true. This explains why 'can have' seems odd. However, some judgments like (CPM) might be true if actions are bodily movements. For example, if I shot someone a few minutes ago, and whether or not my victim will die depends on whether or not I now call a doctor, and I can now call a doctor or not, then I can now determine whether or not my past act will be describable as an act that caused the victim's death, so I can now bring it about that I did or did not kill the victim. Thus, a judgment like (CPM) seems to be true. However, even if judgments like (CPM) are never true, this would support rather than refute my main argument. It is the fact that 'ought' often does operate on past tense sentences to produce a truth but 'can' usually does not operate on past tense sentences to produce a truth that shows that 'ought' does not semantically imply 'can'.

A third possible response is that judgments like (OPM) are never true. However, such a universal claim cannot be justified without begging the question. Furthermore, it often seems natural to say that someone ought to have done something, even if he or she did not do it. We need to be able to make present judgments about what ought or ought not to have been done in the past in order to justify present punishment for past acts or failures to act. The correct interpretation of such judgments cannot have the form of (POM), because judgments like (POM) are true in many cases where punishment is not justified, such as when it was no longer true at the time at which the agent did the act that he or she ought not to do it. Thus, it must be judgments like (OPM) that seem natural and needed.

A final response is that, even if 'ought' is used to make judgments like (OPM), and even if judgments like (OPM) do not semantically imply judgments like (CPM), such judgments use 'ought' in one sense, but there is still a different sense in which 'ought' semantically implies 'can'. A move like this can always be tried to escape arguments against semantic implications. Even if I argue for the obvious conclusion that 'father' does not semantically imply 'married'. an opponent can respond that my arguments show only that one sense of 'father' does not semantically imply 'married', but there is still another sense of 'father' that does semantically imply 'married'. Such moves are not plausible unless there is enough independent justification for distinguishing the supposedly different senses. In the case of 'ought', an opponent might argue that 'ought' is used for different speech acts when it is followed by a past infinitive and when it is followed by a present infinitive. A speaker can advise by saying what the audience ought to do but often cannot advise by saying what the audience ought to have done. However, this difference in possible speech acts does not prove any difference between the senses of 'ought' when it is followed by infinitives of different tenses. In the absence of any other justification for distinguishing such senses of 'ought', it seems simpler and better not to multiply the senses of 'ought'. This leaves no escape from the conclusion that 'ought' does not semantically imply 'can'.

Dartmouth College, Hanover,
New Hampshire 03755, U.S.A.

© WALTER SINNOTT-ARMSTRONG 1985

REFERENCES

[1] H. P. Grice, 'Logic and Conversation', in *The Logic of Grammar*, ed. D. Davidson and G. Harman (Encino, California; Dickenson Publishing Co., 1975), pp. 64–75.
[2] W. Sinnott-Armstrong, '"Ought" Conversationally Implies "Can"', *Philosophical Review*, vol. XCIII, no. 2 (April 1984, pp. 249–61).

DELIBERATING ABOUT THE INEVITABLE

By BRUCE N. WALLER

IF I believe that I shall inevitably perform a certain act — that it is uniquely determined that I shall perform a certain act — can I *deliberate* about what act to perform? A crucial step in Peter van Inwagen's argument for universal belief in free will (*An Essay on Free Will*, Oxford 1983, pp. 153–61) requires a negative answer to that question. I shall argue for the affirmative.

Briefly, van Inwagen's argument: first, 'One cannot deliberate about whether to perform a certain act unless one believes it is possible for one to perform it.' (For van Inwagen, to believe that it is *possible* to perform an act is to believe that one is deciding between or among 'various incompatible courses of action' (p. 154; also pp. 30, 68–9) and that each of those courses of action could actually be selected and put into practice. Also the *possibility* of performing a certain act is not merely a conditional possibility: I *could* have done X *if* I had wanted to, had had different desires, had chosen differently, etc. Van Inwagen rejects this 'conditional analysis of 'could have acted differently' in his superb critique of compatibilism, pp. 114–26.) Two, if one believes determinism to be true, then one believes that one's future acts are inevitable (and that no other acts are really possible). Three, we do deliberate about whether to perform certain acts. Therefore, we cannot believe in determinism.

Van Inwagen's argument is valid: by *modus ponens* on the first and third premisses we derive the consequent of the first premiss, and then by *modus tollens* with the second premiss we validly conclude that no one (at least no one who deliberates) believes in determinism. Furthermore, the second and third premisses are surely true. Thus the vital question is the truth of the first premiss: is it true that 'one cannot deliberate about whether to perform a certain act unless one believes it is possible to perform it'?

Van Inwagen suggests that anyone who doubts such an obvious fact should:

> ... imagine that he is in a room with two doors and that he believes one of the doors to be unlocked and the other to be locked and impassable, though he has no idea which is which; let him then attempt to imagine himself deliberating about which door to leave by (p. 154).

But this example is too simple, its conclusions too obvious. There is no real decision about which door to leave by: the other alternative is obviously (and literally) barred. But in this case deliberation is precluded by the salience – not the inevitability – of the conclusion. In a similar manner, I cannot deliberate whether to walk or (flap my arms and) fly to the tavern; but I *can* deliberate about whether to walk or drive, even if I firmly believe that the decision resulting from that deliberation (as well as the decision to deliberate) is determined (by causes set in inexorable motion long before St Patrick drove the snakes out of Ireland).

The possibility of determined deliberation is clearer in a more complex example: what sort of wager to make on the Grand National Steeplechase. I am a determinist, and I believe that my ultimate choice is completely and uniquely determined (by my learning history, my past patterns of reinforcement, the current influences on me): my bet will be the causal result of those determining factors, and I could not choose other than I shall choose. Will I now find it impossible to deliberate about a number of options, since I believe that only one is really possible? Certainly not. For I must still carry out the determined deliberation process in order to arrive at my betting choice; just as I must still trudge to the betting window, even if I believe that it is determined that I shall arrive at that window. I recognize that my deliberation process – involving my beliefs, memories, hopes, changing knowledge of odds and jockeys and weather conditions – is an essential part of my choosing among various horses. Even though that deliberation process is completely determined it will be no less a process of genuine deliberation and no less important in selecting a fancy (it will still play a causal role in determining my wager). Without that deliberation process (if I instead bet on an appealing name, or my reasoning process were addled by drugs) the final selection would probably be quite different. It is determined *that* I shall deliberate, *how* I shall deliberate, and what the *result* of my deliberation will be; but why should any of that preclude deliberation or lessen its importance?

But (determined) deliberation is not only possible when the course to be taken is unknown; deliberation is even possible in the cases that seem most congenial to van Inwagen's analysis: deliberation is possible even when the *result* of the deliberation is foreknown, and is acknowledged to be determined, fixed, and inevitable. Imagine that computer simulations have been developed to a remarkably sophisticated level, so that an enormously complex simulation takes into account all the influences impinging upon each individual (the effects of interaction between individuals and their social and physical environments, the reasoning processes and emotional patterns of each individual) and projects precisely what each individual will decide and do. (Whether such perfect predictability of rational processes is possible is of course a vexed question. I am only considering the implications of such an obviously deterministic-mechanistic system, not assuming that it is genuinely possible.) This computer simulation is so accurate that I have come to consider it practically infallible, and it would never occur to me to doubt its projections. Suppose further that I am now deliberating my wager for the Grand National Steeplechase, and I am reliably told that my projection reveals that I will bet on Corbiere. At this point I know (or at least I have no doubt) that I shall bet on Corbiere. Is it still possible for me to deliberate about my choice? Certainly. Suppose I had rather fancied Greasepaint, and thought it likely that my deliberations would yield Greasepaint as my betting choice. I am thus surprised by the simulation result (I have no doubt that it *is* the result that I shall deduce), and proceed through the deliberation process in order to see exactly *how* I decide on Corbiere. Thus even if I firmly believe it is not possible for me to do anything other than wager on Corbiere, I can still deliberate about betting on Greasepaint; and van Inwagen's claim '... that no one could deliberate about whether to perform an act that he does not believe it possible for him to perform' (p. 156) is mistaken.

But perhaps that is not sufficient. For van Inwagen elsewhere acknowledges that one *can* deliberate while believing in determinism, at the cost of engaging in contradictory behaviours. Van Inwagen claims that '... to reject free will is to condemn oneself to a life of perpetual logical inconsistency,' that '... anyone who denies the existence of free will must, inevitably, contradict himself with monotonous regularity' (p. 160). This follows from the fact that we must deliberate, combined with '... the fact, for fact it is, that in deliberating one manifests in one's behaviour a belief in one's free will with respect to the act under deliberation' (p. 158). And to specify the belief that is necessarily manifested by deliberation:

> In my view, if someone deliberates about whether to do A or to do B, it follows that his behaviour manifests a belief that it is *possible* for him to

> do A – that he *can* do A, that he has it within his power to do A – and a belief that it is possible for him to do B (p. 155).

So if one denies free will, one might still be able to deliberate; but on pain of contradiction, since one's deliberation necessarily manifests belief in free will.

But why *must* an individual who believes that the result of his/her deliberation is completely determined (and that the resulting choice and behaviour are absolutely fixed) none the less manifest (by deliberating) belief in free will, belief that the deliberation result is genuinely open? The deliberation processes of most people may manifest such beliefs, since most people probably believe in free will and believe that they are deliberating among genuinely open possibilities; but ordinary manifestations are not helpful here. The question is whether it is *possible* to deliberate without manifesting belief in free will. Speaking of illness at one time usually manifested belief in demons; but there was then and is now no inconsistency in speaking of illness without believing in demons. Deliberating may ordinarily manifest belief in free will; but there is no inconsistency in deliberating without believing in free will; and if others should believe that a determinist's deliberation manifests belief in free will, then those others are mistaken – just as a believer in demon possession would be mistaken in concluding that a modern physician's diagnosis of illness manifests belief in demons.

Why does it supposedly follow that deliberation *must* manifest free will? Obviously not because when one observes deliberations one always sees something additional manifested: belief in free will. (Even if one did make such observations, they would not establish that deliberation *must* manifest belief in free will.) Van Inwagen's sole grounds for asserting that deliberation must manifest belief in free will is his claim that the deliberation process *requires* belief in free will, and that therefore anyone who deliberates must manifest such a belief. Thus his claim that belief in free will is *manifested* by deliberation rests on his claim that belief in free will is *required* for deliberation. But if van Inwagen's reason for claiming that deliberation manifests belief in free will is his assertion that belief in free will is necessary for deliberation; then obviously the conclusion that deliberation manifests belief in free will cannot be used to support the claim that belief in free will is a condition of deliberation. Or at least it cannot be so used without making the argument spin in place.

In conclusion: van Inwagen has no independent grounds for maintaining that deliberation must manifest belief in free will; nor does he establish that deliberation requires belief in free will. And the counterexamples and answering arguments indicate that deliberation is compatible with believing all elements of deliberation – including processes of deliberation, results of deliberation, and decisions to deliberate – are completely and uniquely determined.

Or perhaps not. Van Inwagen confidently asserts that:

> ...all philosophers who have thought about deliberation agree on one point: one cannot deliberate about whether to perform a certain act unless one believes it is possible for one to perform it (p. 154).

This gives me pause, for it seems to follow – since I do not agree on that point – that I am not a philosopher; and thus I may have been drawing my meagre salary under false pretences. Thankfully, however, the dilemma has another horn: from the outset I believed that the conclusion I would draw would be (given the determining causes) the only one I could possibly reach; so – if deliberation requires belief in free will – perhaps I have not been deliberating about deliberation at all.[1]

Elon College,
Elon College, North Carolina 27244,
U.S.A.

© BRUCE N. WALLER 1985

---

[1] Thanks to John G. Sullivan, Will Migniuolo and the Editor for valuable comments and suggestions.

IMPECCABILITY REVISITED

By W. R. CARTER

AT the end of his paper 'Impeccability' (ANALYSIS 43.2, March 1983) Thomas V. Morris concludes that the 'traditional theist thus seems to be on quite solid ground in affirming that God is impeccable, or essentially sinless, as well as being all-powerful. If there is any problem about impeccability, or the traditional concept of God incorporating impeccability, it remains to be shown' (p. 112). It is charged that my earlier argument in behalf of inconsistency ('Omnipotence and Sin', ANALYSIS 42.2, March 1982) fails to establish that any occupant of the divine office must be essentially sinless (impeccable). Morris is nonetheless inclined to *accept the thesis that God must be essentially sinless.*[1] His position is that the impeccability thesis is both true (though not demonstrated by my arguments) and innocent of the charge of generating inconsistency.

---

[1] See especially page 44 of Morris' more recent paper 'Properties, Modalities, and God', *The Philosophical Review* XCIII (January 1984).

For reasons set forth below, I believe that the inconsistency charge can be shown to stick. Morris tells us that:

> The problem has been that if God is impeccable, or logically unable to sin, then it seems that there are possible actions for a being with his other properties which he is thus unable to perform, or possible states of affairs which he cannot in any sense actualize. And if this were true, then subject to all the required qualifications, it [impeccability] would be incompatible with God's being omnipotent. However, by drawing attention to a simple and independently plausible distinction, Joshua Hoffman has suggested that any such acts or states of affairs we can describe or conceive are merely conceivable and not at all possible. In a less than Cartesian sense, the God who is impeccable is the ground of all possibility. Our ability to describe situations which involve God's contravening some duty should just remind us of the distinction between conceivability and possibility. They do not coincide. And omnipotence ranges over only what is possible (p. 111).

The general idea behind this is, I think, as follows: since Yahweh (a hypothetical occupant of the divine office) is impeccable, there is no possible state of affairs (world, if you will) in which Yahweh sins, and so no possible state wherein (say) Yahweh murders someone. Accordingly, Yahweh's inability to sin (murder) poses no threat to Yahweh's alleged omnipotence. To be omnipotent one must be capable of doing anything that is possible. Since it is not possible for an impeccable being to sin, the fact that Yahweh cannot sin fails to establish that Yahweh fails to be omnipotent. Perhaps we can *conceive* of (imagine) Yahweh murdering someone; but such a thing is not really possible.

This may sound plausible. But when we look closely I think we find that something is amiss. Let us consider a group of sentient beings, perhaps Martians, who worship a "higher" being they call 'Dennis'. Dennis exists, as most Martians believe, if and only if something occupies a certain 'role' or 'office' (the Dennis-office). The conditions for filling this role are said to be (1) omnipotence, and (2) essential incapacity to solve even the most elementary problems concerning plain geometry. (For some reason Martians believe that anyone who can do plain geometry must have serious moral failings.) Martian atheists, so to speak, argue that it is impossible for anything to occupy the Dennis-role. Any being (they argue) who satisfies requirement (2) thereby fails to satisfy condition (1). So reasoning, atheists conclude that there is no possible world in which Dennis exists. Of course Martian "believers" have a reply to this. Since Dennis is essentially incapable of doing geometry, there is no possible world in which Dennis solves geometry problems. Being omnipotent is a matter of being able to do what is possible. Since it is not possible for Dennis to do geometry, nothing licenses the conclusion that Dennis is not omnipotent. One can, Martian believers remind us, conceive of things that aren't *really possible.* And omnipotence 'ranges over only what is possible'.

To my mind there is something slightly ludicrous about this argument in behalf of Dennis-consistency. Suppose that we were to go along with the argument. Let us substitute in place of (2) condition (3): *logical incapacity to formulate plans and to act intentionally*. To occupy the Dennis-office a being must satisfy both (1) and (3), we are told. The fact that Dennis cannot act intentionally is, we are further told, no impediment to the claim that Dennis is omnipotent. Since it is not possible for any satisfier of (3) to perform such deeds, and since omnipotence ranges only over what *is* possible, it is possible for something that is incapable of opening doors or closing windows to satisfy the conditions definitive of Dennis-hood.

The concept of omnipotence is being stretched a bit here. Indeed it is fair to say that the concept is distorted beyond recognition. No being that is incapable of doing geometry – certainly no being that is incapable of acting intentionally – qualifies as omnipotent. Perhaps we can say that being omnipotent is a matter of being subject to no constraints. But we should not proceed to judge that '... an agent is constrained only when what he is prevented from doing is something that in some possible circumstances or conditions he is able to do'.[2] No being whose range of possibilities fails to include solution of simple geometry problems has any genuine claim to *omnipotence*. The point bodes ill for the claim that Dennis-hood is a consistent concept. And a similar point undercuts the Hoffman-Morris case for consistency when it comes to the concept of a being who is both omnipotent and impeccable (essentially sinless). Such a being not only does not but *cannot* do many things (murder, rape) that ordinary humans can do! Such a being has no claim at all to being judged omnipotent. I conclude that Morris is mistaken when he says that a 'careful explication of the notions of omnipotence and impeccability will allow both to be conditions of deity' (p. 111). Or, if these are conditions for deity, there is no possibility at all that God exists.

Earlier I argued that no occupant of the divine office could *vacate* or *forfeit* office by virtue, so to speak, of moral malfeance. Morris agrees with this, but denies that it follows that an occupant of the divine office must be essentially sinless.[3] Suppose that Yahweh – a hypothetical occupant of the divine office in world W1 – is not essentially sinless. This commits us to rejecting the idea that impeccability is a necessary condition for occupancy of the divine office. For reasons explained above, impeccability has to be rejected in the event that the omnipotence condition is retained.

---

[2] Joshua Hoffman, 'Can God Do Evil?', *Southern Journal of Philosophy* XVII (1979), p. 219.

[3] Suppose that at no time is there a possible future – relative to the way the world is at the time in question – in which Yahweh sins. Given the analysis of (absolutely) essential properties defended by Michael Dummett, it follows that being sinless is an (absolutely) essential property of Yahweh's. See *Frege* (Duckworth, 1973), pp. 130–1. But for reasons set forth in another paper I think that Dummett's analysis of absolutely essential properties should be rejected. Thus I agree with Morris that my earlier malfeance argument failed to show that an occupant of the divine office must be impeccable.

Yahweh only happens to be sinless in W1. So there is a world, W2, in which Yahweh sins. Since moral perfection is a condition for holding the divine office, Yahweh fails to be God in W2. Though Yahweh *is* God (has the attribute of being God) in W1, being God is only a contingent, as not an essential, characteristic of Yahweh. The suggestion that something that is God just *happens to be God* is perhaps a bit unorthodox. But if the argument given above is sound, and impeccability and omnipotence are inconsistent, then perhaps such a view is a necessary 'departure from orthodoxy'.[4] If we follow this line, we are faced with some interesting questions. Does the divine office *have* an occupant in W2? Suppose the answer is that it does not. In that case, we must reject the common assumption (which perhaps underlies many *a priori* arguments for God's existence) that it follows from the possibility that God exists that it is necessary that God exists. Atheists may (well) rejoice in the failure of this assumption. Granting the consistency of the concept of God, they may see no reason for judging that 'our' world is a member of the proper subset of possible worlds in which God exists.

Suppose, on the other hand, that the divine office has an occupant in W2. Let the term 'Alice' designate the being that is God in W2. Presumably necessary existence is a condition for occupying the divine office. So it appears that Alice necessarily exists, and so exists in W1. Since Yahweh is God in W1, this means that there exists in W1 a being (Alice) who fails to be God and yet could be God. Given our present asssumptions, a similar argument can be constructed for judging that if God exists in 'our' world then there exists in our world some being who fails to be God but who might be (might have been) God. This is an unorthodox view, to say the least. I suspect that some such unorthodoxy is the price that must be paid for divine consistency.

North Carolina State University,
P.O. Box 5688,
Raleigh, N.C. 27650, U.S.A.

© W. R. CARTER 1985

---

[4] On p. 111 Morris charges that Nelson Pike's proposal for resolving the putative inconsistency involves 'a wholly unnecessary departure from orthodoxy'. The thesis of the present paper is that considerable departures from orthodoxy are required, if consistency is to be achieved.

HAIRIER THAN PUTNAM THOUGHT

By STEPHEN READ *and* CRISPIN WRIGHT

FROM one point of view, it may seem that Sorites Paradoxes are a consequence of the very nature of vagueness. If 'bald', for example, lacks sharp boundaries, that can only mean, it appears, that there is no determinate number of hairs, n, such that any man with n hairs on his head is bald while some man with $n + 1$ hairs on his head is not. Accordingly it seems we must deny

(1) $(\exists n)(\mathrm{BM}(n) \,\&\, {\sim}\mathrm{BM}(n + 1))$

(where '$\mathrm{BM}(n)$' is short for 'Any man with n hairs on his head is bald'). Classically this denial is equivalent to assertion of

(2) $(n)(\mathrm{BM}(n) \rightarrow \mathrm{BM}(n + 1))$,

which generates a Sorites paradox. Conjoined, for example, with the undeniable

(3) $\mathrm{BM}(0)$,

(2) will lead, either by successive applications of universal instantiation and detachment, or by induction, to

(4) $\mathrm{BM}(50{,}000)$

which is clearly false. If we essay to treat this result as a *reductio* of (2), we wind up with

(5) ${\sim}(n)(\mathrm{BM}(n) \rightarrow \mathrm{BM}(n + 1))$

which is classically equivalent to (1). Hence, it appears, if any predicate is vague, it must be Sorites-generating (or *tolerant*); and the option of treating Sorites paradoxes as a *reductio* of their major premises is available only at the cost of denying the vagueness of the predicates involved – which is hardly less paradoxical.

In 'Vagueness and Alternative Logic' (*Realism and Reason*, Cambridge 1983, pp. 271–86, especially 285–6), Hilary Putnam puts forward a suggestion for a formal treatment of the logic of vagueness. His leading thought, although not presented in the context of exactly the foregoing perspective on the problem, is in effect that the equivalence of (5) with (1) is not intuitionistically valid. May not a suitable treatment for vague predicates be available by treating them in the manner in which intuitionistic logic treats predicates in number theory which are not effectively decidable? The prospect is thereby opened of treating the paradox as indeed a *reductio* of (2), so accepting (5), without commitment to (1) with its paradoxical imputation of precision to 'BM(x)'.

Putnam admits that, at the time of writing, he had not thought this idea through. What will already be apparent to the alert reader is that, in order to disclose serious difficulties for the proposal, Putnam would not have had to think very far. For if the denial of (1) is indeed a satisfactory expression of the vagueness of '$BM(x)$', intuitionistic logic provides all the materials needed for the paradox. This may be seen in either of two ways. The denial of (1) is intuitionistically equivalent both to

(6) $(n)(BM(n) \rightarrow \sim\sim BM(n + 1))$

and to

(7) $(n)[\sim BM(n + 1) \rightarrow \sim BM(n)]$

Now appropriately many steps of contraposition (which is of course intuitionistically valid) will obtain from (6) any principle of the form

(8) $(n)[\sim^{k} BM(n) \rightarrow \sim^{k+2} BM(n + 1)]$

(where '$\sim^{k} BM(n)$' is to indicate the result of prefixing '$BM(n)$' by some even number, k, of occurrences of '$\sim$'). By appropriate successive appeal, then, to the members of the series of such principles, there will be no difficulty (in principle) in advancing from (3) and (6) as initial premises to, say

(9) $\sim^{100,000} BM(50,000)$.

But the admitted falsity of (4) yields

(10) $\sim BM(50,000)$

which in turn entails

(11) $\sim(\sim^{99,998} BM(50,000))$

of which (9) is the negation. Alternatively, and more simply, (7) will generate a Sorites paradox back-to-front, as it were, starting from (10) and leading to the denial of (3). So the paradox is with us yet.

Although Putnam's proposal is unsatisfactory, its failure is instructive: if the *denial* of principles like (1) is considered to follow from the vagueness of the relevant expressions, then no logic which, like intuitionistic logic, endorses *reductio ad absurdum* and the classical handling of the quantifiers and conjunction can prevent the derivation of Sorites paradoxes. This is evident from the starkest form of the paradox, a back-to-front version each phase of which proceeds as follows:

i	(i)	$\sim(\exists n)[BM(n) \,\&\, \sim BM(n + 1)]$	Assumption
ii	(ii)	$\sim BM(k - 1)$	Assumption
iii	(iii)	$BM(k)$	Assumption
ii, iii	(iv)	$BM(k) \,\&\, \sim BM(k + 1)$	(ii), (iii), &I
ii, iii	(v)	$(\exists n)[BM(n) \,\&\, \sim BM(n + 1)]$	(iv), EI
i, ii	(vi)	$\sim BM(k)$	(iii), (i)/(v), RAA

Certain Relevant Logics do indeed, of course, modify the appropriate classical rules.[1] But intuitionistic logic doesn't. Moreover, it is difficult to see how an intuitionistic semantics for negation, appropriately generalized to handle non-mathematical contexts, could fail to endorse the negation of (1). For, intuitionistically, one is entitled to assert the negation of a statement just in case one has verified that no constructive proof of that statement is in principle possible; and to recognize the vagueness of 'BM(x)' is surely exactly to recognize that all possibility of a *constructive* verification of (1) can be discounted.

Suppose it agreed that it cannot be a satisfactory response to the paradox to admit the inconsistency of vague expressions. What we have seen is that the *denial* of (1) cannot be, for anyone with classical or intuitionistic sympathies, a satisfactory way of expressing the vagueness of 'BM(x)'; and, as noted, the *assertion* of (1) merely flies in the face of its vagueness. It follows that a satisfactory account of the matter must render it coherent to hold that (1) may correctly be neither asserted nor denied; although we may take the view that its negation, and (2), may, and indeed must, be denied, since refuted by the paradox. Consequently the unrestricted validity of Double Negation Elimination for statements containing vague expressions is indeed in doubt, and Putnam's proposal has point to that extent. But only to that extent. For, as noted, the motivation for intuitionistic-type restrictions in the logic of vague expressions cannot closely resemble orthodox intuitionistic semantics if we are to avoid denial of (1); and, further, until a way has been found of properly acknowledging vagueness while avoiding such a denial, recourse to intuitionistic logic is no help.[2]

Department of Logic and Metaphysics,
The University, St Andrews,
Fife KY16 9AL

© STEPHEN READ *and*
CRISPIN WRIGHT 1985

---

[1] The central tenet of relevant logic is that the mere lack of a counterexample does not guarantee an entailment. Hence, (i) will not, without qualification, entitle us to pass from (ii) to (vi). Note that this move would be an instance of *modus ponendo tollens*, classically and relevantly equivalent to *modus tollendo ponens*, i.e. Disjunctive Syllogism, which is, familiarly, relevantly invalid. What *in detail* a 'relevantist' will demur at in the reasoning is another matter.

[2] We are indebted to Peter Clark and John Slaney for their contributions to the discussion which suggested this note.

McGINN'S REPLY TO WRIGHT'S REPLY TO BENACERRAF

By ALLEN HAZEN

IN his classic article 'What Numbers Could Not Be' (*Philosophical Review*, 1965; reprinted in the 1983 edition of Benacerraf and Putnam's *Philosophy of Mathematics*), Paul Benacerraf argued against 'logicism', on the grounds that the impossibility of specifying one of the many possible reductions of arithmetic to set theory made it hopelessly artificial to claim that numbers are sets, or indeed objects of any specifiable kind. In his stimulating recent book *Frege's Conception of Numbers as Objects*, Crispin Wright has urged, as against Benacerraf, that the problem he presented for the logicist is simply a special case of Quine's indeterminacy of translation, and so doesn't show that regarding numbers as objects is any less defensible than regarding rabbits as objects. Now (ANALYSIS 44.2, March 1984) Marie McGinn has stepped in and claimed that Benacerraf's problem is really different from the Quinean ones after all.

I would like to suggest two criticisms of McGinn. First, the weight of authority. In his contribution to the 1969 Quine festschrift *Words and Objections*, Gilbert Harman suggests that the multiple reductions of arithmetic to set theory be taken as a paradigmatic case of Quinean indeterminacy, and in his reply at the end of the book Quine himself says 'Right on, Gil, it's a great example and I wish I'd thought of it myself' (or words to that effect).

Second, McGinn's attempt to distinguish between the Benacerrafian and Quinean indeterminacies makes essential appeal to what I think are special features of the most familiar examples. She points out that 'correcting for' such deviant translations of *gavagai* as *rabbit stage* and *undetached rabbit part* involves mucking about with our translations of (what we would ordinarily take as) singular and plural forms and the identity predicate. With the various reductions of arithmetic – the Zermelo numbers and the Von Neumann numbers are the usual examples given – there is no such involvement of the structural, or topic neutral, elements of the language. Only the specifically arithmetical vocabulary is affected: the different choices as to what sets to identify the numbers with entail different translations of such items as *number*, *sum*, *product*, and the names for the individual numbers. (To be quite Quinean, of course, we should eliminate singular terms, and construe the numerals, not as names, but as special predicates on the model of *pegasizes*: *is-zero*, *is-one*, and so forth.) The identity predicate between terms or variables for numbers gets translated by the identity predicate. From which she concludes that Benacerraf's problem is a special one about numbers: there is something wrong with the notion of numbers as objects over and above the general Quinean puzzles about translation and reference.

But *gavagai* has other translations which allow the same translations for the 'auxiliaries of objective reference' as work when it is translated *rabbit. Surface of a rabbit's body* and *volume of space wholly containing and wholly occupied by a rabbit* are two that come to mind; the usual examples show that different translation manuals will sometimes differ on the structural particles, but fixing our translations of the particles can't entirely cure the inscrutability of reference. On the other side, one of the most interesting ways of reducing arithmetic to set theory (I have been told that Zermelo himself thought it closer to capturing the meaning of arithmetic expressions, though technically far less convenient, than his official reduction) does require that arithmetic identity be translated by something other than set theoretic identity. On this scheme, quantification over (natural) numbers is construed as quantification over arbitrary (finite) sets, and the identity predicate, when standing between numerical terms or variables, is translated *is in one-one correspondence with.*

I think the moral is that Benacerraf's argument refutes, not Platonism in general, but at most a very special brand of Platonism, one that holds that mathematics is about a definite domain of mathematical objects, the sets, and that, in order for arithmetic to count as a branch of mathematics, some one among the many interpretations of number theory in set theory must be an analytic reduction. Such a view is indefensible, I think, and Benacerraf's objection, even though it is closely related to Quine's indeterminacies, helps to show this. A more liberal sort of Platonism, however, is also possible. In this view, mathematics has a fully objective, or 'real', subject matter, but this subject matter can be 'carved up' into objects in various ways. (Compare: an old-fashioned geometer who thought Euclidean geometry was about physical space, but who was sophisticated enough to realize that space could be conceived of either as an assemblage of points, lines being reduced to pairs of points, or as an assemblage of lines, points being reduced to pairs of intersecting lines.) On such a view the possibility of interpreting number theory in set theory can still be of great importance: it is evidence that the set-theoretic way of conceiving mathematical reality is complete, in the sense in which the conception of the Euclidean plane as a domain of points related by a three-place collinearity and a four-place equidistance relation is complete. This does not, however, require that any such interpretation be an analytic reduction. Finally, even if the set-theoretic conception *is* complete, in the sense that every fact about the subject matter of mathematics is representable as a fact about sets and their relations of membership, it need not, on this view, be unique. A category-theoretic conceptualization, for example, might be equally complete. And a conception of numbers as objects, unreduced, though it will not be complete, will also be a perfectly legitimate conceptualization of part of the truth about the subject matter of mathematics.

NOTE: The 'liberal Platonism' I have described in the last paragraph is, I think, consonant with the views of Bernays, who wrote that '[m]athematics ... can be regarded as the theoretical phenomenology of structures'. ('Comments on Ludwig Wittgenstein's *Remarks on the Foundation of Mathematics*', in the 1964 edition of Benacerraf and Putnam.) It is also in the spirit of the remark, attributed to Kreisel, that the real issue of realism in the philosophy of mathematics is not over the existence of mathematical objects but over the objectivity of mathematical truth. More recently, Mark Wilson ('The Double Standard in Ontology', *Philosophical Studies*, 39, 1981) has argued the compatibility, with a broadly realist view, of various ways of 'carving reality up' into objects, discussing (a more sophisticated version of) the example about lines and points. Indeed, even some of Benacerraf's own remarks, about arithmetic being the study of a certain 'abstract stucture', can be interpreted in a sense consistent with 'liberal Platonism'.

Melbourne University,
Parkville, VIC 3052, Australia

© ALLEN HAZEN 1985

WHY KANT COULDN'T BE AN ANTI-REALIST

By A. J. CLARK

WAS Kant the first semantic anti-realist? There is a strain of thought in modern philosophy which would argue that he was. Hilary Putnam suggests that we read Kant as claiming that:

> The only sort of truth we can have an idea of, or use for, is assertability (by creatures with our rational natures) under optimal conditions. (*Realism and Reason*, Cambridge 1983, p. 210)

Carl Posy, too, proposes to 'interpret Kant's idealism along the lines of a modern, non-reductive anti-realism' ('The language of appearances and things in themselves', *Synthese* 47, 1981, p. 313). Whatever the virtues of such discussions there is simply no way (or so I shall argue) that Kant himself could consistently adopt an anti-realist stance which assimilates *all* grasp of meaning to grasp of conditions of warranted assertability. For to do so would be to abandon the vital (but neglected) distinction between the *real* possibility of a concept and its *mere* possibility. To see the importance of this distinction, let us consider the classic objections to Kant's transcendental idealism levelled by Peter Strawson in *The Bounds of Sense* (Methuen 1966).

Strawson objects to Kant's use of the transcendental object. This use, Strawson claims, contravenes Kant's own principle of significance which states that all meaningful concepts must be such as can be given in a possible experience. Thus consider a typical Kantian claim about noumena, e.g.:

> We are unable to comprehend how ... noumena can be and the domain that lies out beyond the sphere of Appearance is for us empty. (I. Kant, *Critique of Pure Reason,* A 255)

And consider also a typical Kantian claim about understanding, e.g.

> The understanding can never transcend those limits of sensibility within which alone objects can be given to us. (op. cit., A 247)

The question which Strawson asks is why claims of the latter kind (which are responsible for the rush to assimilate Kant to the modern anti-realist) do not rule out claims of the former kind. As Strawson puts it, are the claims about noumena not claims of which we *can* comprehend the impossibility (by virtue of the principle of significance expressed in passages such as A 247) rather than, as Kant insists, claims the possibility of which we cannot comprehend?

Perhaps. But what I want to show is that Kant himself had a way of circumventing this difficulty, at least to his own satisfaction. This way out, I suggest, involves apparatus unavailable to the modern anti-realist. So if Kant *were* the first semantic anti-realist, his account would be incoherent in just the way Strawson urges. Kant's escape relies on a distinction between grasping the *real possibility* of a concept or assertion and grasping that it is, at least, not *impossible* that the conditions it describes should obtain. This distinction allows Kant to maintain the bare concept of a noumenal realm in the face of what Strawson took to be his account of what is involved in grasping the meaning of a claim. For the principle of significance, as Kant understood it, may best be seen as describing *not* what is involved in grasping the meaning of a claim *per se* but what is involved in grasping the meaning of a claim that such and such a state of affairs is *really possible.*

Kant certainly had the necessary apparatus to engage in such a manoeuvre. He writes that

> To say that the not-being of a thing does not contradict itself ... is very far from being sufficient for real possibility. (op. cit., A 244)

Real possibility, he asserts, requires that the 'bare concept be made sensible'. And what this involves is 'the possibility of giving it an object to which it may be applied' (A 240, A 239). This, of course, is precisely what is lacking in the case of claims about the existence of a noumenal realm. So Kant consistently concludes that:

> It is ... an open question whether the notion of a noumenon be not a mere form of a concept (op. cit., A 253)

But (and here is the trick) it is only the *bare concept* which Kant requires. We grasp the bare concept when we grasp that the idea of a noumenal reality is not a self-contradictory one. It is thus that we can *think* objects as things-in-themselves even though we cannot *know* (= grasp the real possibility of) them as such (B xxvi). And this, of course, is sufficient for Kant's purposes. For all he seeks (despite those mistaken theories which see him as arguing that noumena can be known to affect the senses and cause appearances) is to make room for *faith* in an extra-phenomenal reality. This can be done *just so long as* we can grasp the meaning of the claim that such a reality is possible (= not impossible) even though it is not really possible, i.e. could not be given in experience.

The anti-realist is entitled to no such distinction. For him, *all* grasp of meaning comes down to a grasp of conditions of warranted assertability. And the warrants are always such as could, in principle, be given in experience. There is no room in such a system for the kind of bare or negative grasp which Kant felt we had of a proposition in terms of its mere possibility. In other words, for the anti-realist, all possibility is real possibility. For to understand what some claim of the form 'possibly P' means is, in part, to grasp the conditions which, were they to obtain, we could *recognize* as warranting the assertion of P. Which is, in turn, sufficient to show that P is, in Kant's terminology, really possible. As Kant himself puts it, if any concept 'can in any fashion verify itself' then it shows 'its real possibility' (B 303). The anti-realist, then, by embracing a broadly verificationist semantics embraces a semantics of real possibility alone. Such a semantics leaves no room for the manoeuvres which, as far as I can see, provide Kant with his only means of securing the critical edifice against Strawson-type objections. Being charitable, I conclude that Kant was not, after all, the first anti-realist.

University of Stirling,
Stirling, Scotland FK9 4LA

© A. J. CLARK 1985

ARE THERE SET THEORETIC POSSIBLE WORLDS?

By SELMER BRINGSJORD

THERE are of course, many conceptions of (possible) worlds currently in fashion. In this note I show that *one* of these conceptions (viz., one which identifies worlds with sets of states of affairs or propositions) is formally incoherent. (My proof is somewhat similar to a paradox Martin Davies presents in his *Meaning, Quantification, Necessity* (Routledge and Kegan Paul, 1981), Appendix 9. However, unlike Davies, I make no use of the rather slippery notion of *thinking a proposition.* And, though Davies perhaps shows that there is no such thing as the set of all worlds, I think I show that there is simply no such thing as a world, in the set theoretic sense.)

The set-theoretic account I attack herein is the following (it encapsulates a synthesis of those set-theoretic construals of possible worlds championed by Robert Adams and Alvin Plantinga):

(D1) w is a world =df. w is a set of states of affairs such that (i) for every state of affairs p: either $p \in w$ or $\sim p \in w$; and (ii) the members of w are compossible.

The proof is as follows. Take any world w; and assume that the cardinal number of w is k. Consider, next, the power set of w, which we will denote by 'P(w)'. The cardinality of P(w), by Cantor's theorem, will be 2^k; and $2^k > k$. But for each element $e_i \in P(w)$ there will be a corresponding state of affairs (or, if you like, a proposition) having the form *e_i's being a set.* (Any old state of affairs will do here: *e_i's being an element of P(w)*, etc.) And it follows from (D1) that for each of these states of affairs, either it or its negation will be an element of w. Hence the cardinal number of w is at least 2^k. But we started with the assumption that the cardinal number of w is k.[1]

Brown University,
Providence, Rhode Island 02912,
U.S.A.

© SELMER BRINGSJORD 1985

---

[1] I am indebted to Roderick Chisholm, Michael Zimmerman, Atli Hardarson, and an anonymous referee.

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol 62, No. 4; December 1984

EDITOR: BRIAN ELLIS

ARTICLES

John Heil	Reliability and Epistemic Merit
F.C. White	The Scope of Knowledge in Republic V
Michael B. Burke	Hume and Edwards on 'Why Is There Something Rather Than Nothing?'
Reinhardt Grossman	Nonexistent Objects Versus Definite Descriptions
Rolin Church	Popper's 'World 3' and the Problem of the Printed Line
Reed Richter	Rationality Revisited
Richard Creath	Smart, Salmon, and Scientific Realism

REVIEW ARTICLE

Knud Haakonssen	Hume: Realist and Sceptic

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$16 (US$18 or Stg£9): Institutions A$32, (US$36 or Stg£17). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol 63, No. 1; March 1985

EDITOR: BRIAN ELLIS

ARTICLES

Chris Mortensen	The Limits of Change
Les Holborow	Benn, Mackie and Basic Rights
J. Christopher Maloney	About Being a Bat
Alan Brinton	A Rhetorical View of the *Ad Hominem*
Bruce Langtry	The Maximin Rule Argument for Rawls's Principles of Justice

DISCUSSION

Eddy M. Zemach	Speaking of Belief
George N. Schlesinger	Inaccessible Routes to the Problem of Privileged Access

CRITICAL NOTICE

Ian Hinckfuss	*The Case for Idealism* by John Foster

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$17 (US$18 or Stg£10); Institutions A$34, (US$36 or Stg£18). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

Personal Identity

SYDNEY SHOEMAKER and RICHARD SWINBURNE

What does it mean to say that this person at this time is 'the same' as that person at an earlier time? If the brain is damaged or the memory lost, how far does a person's identity continue? In this book two eminent philosophers develop very different approaches to the problem.

Great Debates in Philosophy
164 pages, hardback **£15.00** (0 631 13208 2)
paperback **£5.50** (0 631 13432 8)

Consciousness and Causality

D.M. ARMSTRONG and NORMAN MALCOLM

How are the objects of consciousness related to consciousness? Can consciousness be defined as an awareness of inner mental states. Can one perceive consciousness in oneself? Can states of consciousness be identified as brain states? David Armstrong and Norman Malcolm here present opposing views on these and related questions in the philosophy of mind.

Great Debates in Philosophy
228 pages, hardback **£17.50** (0 631 13212 0)
paperback **£6.50** (0 631 13433 6)

From Knowledge to Wisdom

A Revolution in the Aims and Methods of Science

NICHOLAS MAXWELL

In this wide-ranging yet closely argued book, the author argues forcefully the need for a comprehensive revolution in the aims and methods of science. He urges a new kind of enquiry which would have as its fundamental aim the enhancement of personal and social wisdom.
320 pages, **£19.50** (0 631 13602 9)

But is it Art?

B.R. TILGHMAN

The question of how we identify, describe and evaluate works of art has exercised philosophers through the ages. In this lucid, witty and thought-provoking book, Professor Tilghman explores the possibility or impossibility of constructing a proper philosophical theory of art.
208 pages, **£15.00** (0 631 13663 0)

Basil Blackwell

Morality and Conflict

STUART HAMPSHIRE

'Characteristically authoritative . . . Stuart Hampshire compels us to think. A single review cannot do justice to its richness, nor indeed to the moving seriousness with which it is written.' *The Listener*
'Important and influential essays' *Times Higher Education Supplement*
186 pages, hardback **£16.50** (0 631 13336 4)
paperback **£6.50** (0 631 13982 X)

Realism and Truth

MICHAEL DEVITT

This book argues that realism is essentially a metaphysical issue. Drawing on the correspondence theory of truth, Michael Devitt develops a strong defence of realism, which leads him to advance a naturalized approach to semantics and epistemology. In the discussion which ensues he examines and rejects in turn the views of most influential critics of realism.
264 pages, **£19.50** (0 631 13535 9)

Women, Reason and Nature

Some Philosophical Problems with Feminism

CAROL McMILLAN

'It is good to have a serious book on feminist thought.'
Ann Oakley, *British Book News*
Carol McMillan argues an original and provocative case: that feminism, far from presenting a radical challenge to sexism, actually shares its most fundamental and mistaken assumptions about women's place in society.
176 pages, hardback **£16.50** (0 631 12496 9)
paperback **£5.95** (0 631 13894 3)

Dialectic of Nihilism

GILLIAN ROSE

This book fundamentally challenges the radical credentials of post-structuralism. Though Derrida, Foucault and Deleuze claim to have 'deconstructed' metaphysics, Gillian Rose shows that their work has much in common with previous attempts to 'end' the metaphysical tradition.
240 pages, hardback **£22.50** (0 631 13191 4)
paperback **£8.50** (0 631 13708 4)

Basil Blackwell

THEORIA SEGUNDA EPOCA

REVIEW OF THEORY, HISTORY AND FUNDAMENTALS OF SCIENCE.
Department of Logic, Basque Country University, San Sebastián, Spain.

The Spanish philosophical journal THEORIA (Madrid, 1952-1956), suspended in 1956 for political reasons, is now preparing its reappearance for the beginning of 1985. THEORIA, Segunda Epoca continues the traditions of quality and diversity of the old Spanish review THEORIA. Like its predecessor, it will focus on the theories, the history, the methodology and the foundations of the sciences, whether formal, natural or social sciences. Yet it will be philosophically comprehensive and will be open to all the philosophical problems, especially those in the philosophy of language, the philosophy of mind and the theory of knowledge; it will also be especially open to the theoretical and technical problems in the new interdisciplinary research, e.g., cognitive science, artificial intelligence, computational linguistics and applications of information theory to any type of science.

The General Editor is Professor Miguel Sánchez-Mazas, the two associate Editors are: Professor Miguel Ángel Quintanilla and Professor Víctor Sánchez de Zavala, and the two editorial Sercretaries are: Professor Fernando Broncano and Professor Javier Echeverría. All correspondence from within Spain should be sent to: Theoria, Departamento de Lógica, Apartado 1594, 20030 San Sebastián, Spain. All other, foreign correspondence should be sent to: Theoria, Asociación Cultural España, Case 2730, 1211 Geneva 2, Switzerland.

Subscription: Yearly volume, US$20.00; single issue, US$8.00.

Journal for the Theory of Social Behaviour

General Editors: C.W. Smith, P. Collett,
and S. Moscovici

Necessary and exciting reading for sociologists interested in the theoretical underpinnings of sociology, and the interrelationship of sociological theory with modern developments in the philosophy of the social sciences and social psychology. A truly interdisciplinary journal, but one with a thoroughgoing sociological sensitivity.

Some recent articles of interest to sociologists include: A special issue on "the duality of structure and structuration", with pieces by Roy Bhaskar, Anthony Giddens, Rom Harre and Allan Pred, among others An article by Douglas Porpora, "On the prospects for a nomothetic theory of social structure", with responses by Peter Blau, Jonathan Turner, Norbert Wiley, Bruce Mayhew and Jefferey Isaac.

1985 subscriptions:
Institutions £32.75 (inland), £40.95 (overseas), $81.00 (USA), $102.50 (Canada);
Individuals £14.50 (inland), £18.15 (overseas), $29.75 (USA), $39.95 (Canada).
Single copies are available, please write for prices.

Orders to: JOURNALS DEPARTMENT, **BASIL BLACKWELL,**
108 Cowley Road, Oxford OX4 1JF, England.

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; Andrew Harrison, R. H. Holland, J. H. McDowell, D. H. Mellor, M. A. Stewart, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices are as follows:

MAY 1985

	U.K.	Overseas	U.S.A.	Canada
Institutions	£11.00	£13.50	$33.00	$39.50
Individuals	£7.45	£9.25	$21.00	$25.00

BACK ISSUES. Volumes 1-37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30-32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

U.S. MAILING. Second class postage paid at New York. Postmaster: Send address corrections to: Analysis: c/o Expediters of the Printed Word Ltd, 515 Madison Avenue, New York, NY10022 (U.S. Mailing Agent).

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

Vol. 45 No. 2 (New Series No. 206) March 198

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CALCUTTA UNIVERSITY CENTRAL LIBRARY

CONTENTS

Nominalism and abstract entities ALLEN HAZEN
Rejoinder to Tenant ALAN WEIR
Buckner quoting Goldstein and Davidson on quotation J. VAN BRACKEL
Tarski a relativist? HARVEY SIEGEL
Sainsbury on denying a Fregean conclusion STIG ALSTRUP RASMUSSEN
The only x and y principle H. W. NOONAN
Some radical consequences of Geach's logical theories JAMES CAIN
Lemmon on logical relations RALPH KENNEDY
'If A and B, then A' E. J. LOWE
Simultaneous causation and causal chains DOUGLAS EHRING
Losing track of the sceptic ANTHONY L. BRUECKNER
Derek Parfit and Greta Garbo GEOFFREY MADELL
The umpire's dilemma COLIN RADFORD
Reasons, dilemmas and the logic of 'ought' CASEY SWANK
A transparent case for subjectivism C. L. HARDIN
Veridicality: more on Searle ALAN MILLAR
What is a dictator? LESLIE GREEN

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

Philosophical Investigations

Edited by **D.Z. Phillips**

Vol. 8 No. 2 April 1985

CONTENTS

Articles:
The Metaphysics of Wittgenstein's *On Certainty* JOHN W. COOK

Pragmatic Rationality FREDERICK L. WILL

Discussion:
Equality and Length E. W. VAN STEENBURGH

Reviews:
Martin Hollis and Steven Lukes (eds.), *Rational and Relativism* LARS HERTZBERG

Kjell S. Johannessen and Tore Nordenstam (eds.), *Wittgenstein – Aesthetics and Transcendental Philosophy* ALLAN JANIK

Books Received

Published quarterly: January, April, July and October
Subscriptions to Volume 8, 1985
Institutions: *£28.00 (UK) £35.00 (Overseas) $65.55 (US)*
$80.30 (Canada)
Individuals: *£12.95 (UK) £16.20 (Overseas) $30.35 (US)*
$37.20 (Canada)

Orders, with remittance, or sample copy requests, should be sent to:
Iris Taylor, Journals Department, Basil Blackwell,
108 Cowley Road, Oxford OX4 1JF, England

Basil Blackwell

The Philosophical Quarterly

Vol.35 No.138 January 1985

Edited in the University of St. Andrews by an Editorial Board
Chairman of the Board of Editors: LESLIE STEVENSON
Executive Editor: GORDON GRAHAM *Reviews Editor:* ANTHONY ELLIS

Socrates' Disavowal of Knowledge	*Gregory Vlastos*
Spatial and Temporal Parts	*J. Butterfield*
Hume's Naturalism – 'Proof' and Practice	*M.J. Ferreira*
Meaning and the Idol of Origins	*Christopher Cherry*
DISCUSSIONS	
Full Humans and Empty Morality	*John Harris*
Dependence, Reliance and Abortion	*Mark Strasser*
Praying about the Past	*Geoffrey Brown*
Clark on God's Law and Morality	*John Chandler*
How to Navigate the River of Time	*George N. Schlesinger*
A Reply to Schlesinger	*L. Nathan Oaklander*

Subscriptions for 1985
Institutions: £26.00 (UK) £32.50 (overseas) $57.50 (US) $71.50 (Canada)
Individuals: £10.25 (UK) £12.50 (overseas) $26.00 (US) $30.00 (Canada)

Orders, remittances and enquiries should be addressed to: Journals Department,
Basil Blackwell Publisher Ltd., 108 Cowley Road, Oxford OX4 1JF, England.

Philosophical Books

Philosophical Books *aims to provide prompt, scholarly reviews of professional books in philosophy. Besides conventional reviews in both contemporary philosophy and the history of philosophy, it now includes regular Discussion features, where the authors of selected titles have the opportunity to reply to reviewers. There are occasional subject surveys and issues devoted to a single theme.*

Books for review should be sent to the Department of Philosophy, Birmingham University, PO Box 363, Birmingham B15 2TT

Published quarterly: January, April, July and October
Subscriptions to Volume 26 (1985)

Institutions: *£23.50 (UK) £29.50 (overseas) $55.50 (US) $68.00 (Canada)*
Individuals: *£11.95 (UK) £14.95 (overseas) $28.00 (US) $35.00 (Canada)*

Orders, with remittance, or requests for specimen copies and further details should be sent to: Iris Taylor, Journals Department, Basil Blackwell, 108 Cowley Road, Oxford, OX4 1JF, England.

BASIL BLACKWELL · OXFORD

NOMINALISM AND ABSTRACT ENTITIES

By ALLEN HAZEN

PLATONISTS believe in abstract entities, nominalists don't. Much of mathematics, at least as it is ordinarily formulated, presupposes abstract entities (by quantifying over them). An important school, led by Quine and Putnam, holds that the debates between nominalism and Platonism in the philosophy of mathematics are to be adjudicated by observing the (abstract) ontological presuppositions of those parts of mathematics required by natural science. Hartry Field, accepting that school's position as regards philosophical method, has sought to defend nominalism by reformulating (part of) physics in such a way as to attribute all the mathematically significant structure to such putatively concrete entities as mereological sums of space-time points, while not restricting the mathematical complexity of its theories: theories whose usual formulations have seemed to more constructivistically minded philosophers to be typically Platonistic. Somewhere, two very different philosophical questions have crossed paths.

Alan McMichael ('Why Physics can't be Nominalized', ANALYSIS 44.2, March 1984, pp. 72–8) has recently started to disentangle them. Suppose that Field's program can be carried out successfully when extended to the whole of physics, so that every apparent appeal to the properties of a domain of mathematical entities is replaced by postulates attributing corresponding structural properties to a domain of physical, and hence concrete, objects. (As the mathematical sophistication of Field's own work suggests, this is not a trivial supposition, but it has some plausibility. After all, the application of mathematics in physics quite generally proceeds by attributing to some aspect of the physical phenomena to be explained a structure parallel to that of the mathematical objects whose theory is to be applied.) Even on that supposition, McMichael argues, a scientifically minded philosopher would have occasion to quantify over (and so presuppose the existence of) certain abstract entities. Science isn't just *done*; it is also commented upon. We will want to be able to say, for example, that the scientific theories of our time are not the last word (or, should we regain the last century's self-confidence, that they are). As one aspect of this, we might wish to claim, or at least entertain the speculation, that not only do our theories not suffice to predict everything, they do not even suffice to predict everything in principle predictable: that there are true laws which are not entailed by our current theories, and which, in conjunction with appropriate descriptions of the initial state of a system, will support predictions that could not be made appealing only to consequences of current theory. To say this, however, is to quantify existentially over propositions (and, incidentally, to help ourselves to be problematic distinction between laws and 'mere' truths).

One could try to weasel out by claiming that the quantification need only be over sentences, and so can be analysed as substitutional, and so (perhaps), ontologically noncommittal. McMichael forestalls this ploy by concentrating on a deeper sort of incompleteness one might wish to attribute to a theory. Perhaps there are phenomena as yet undiscovered, effects for which our current language has no words. Perhaps, even, these unknown phenomena are relevant to the prediction of the sorts of events we *are* able to describe. (Think of hidden variable theories.) But in that case there is a property of physical things, a natural kind, for which our language has no word: and we have quantified over properties in a way which (as Charles Parsons has pointed out) is not, at least in any straightforward way, analysable as substitutional. McMichael discusses several convincing examples.

No matter how physical theories themselves are reformulable, then, a scientifically minded philosopher will quantify over abstract entities in discussing them: even if natural science itself can be made austerely nominalistic, the philosophy of science cannot. Is this conclusion a comeback for Platonism in the philosophy of mathematics? I would like to argue that it is not. The requirements for quantification over propositions and properties McMichael has noted in the philosophy of science can be met with a weak, predicative second order logic. (Indeed, something like Russell's ramified type theory, without the axioms of extensionality or reducibility, would seem better suited than impredicative systems for those applications in the philosophy of science where the distinction between natural kinds and arbitrary groupings is important.) In contrast, the set-theoretic methods classed as Platonistic by most philosophers of mathematics are immeasurably stronger: even the system called 'Classical Analysis' amounts to full, impredicative, second order logic with the natural numbers taken as individuals, and it constitutes but a small fragment of set theory. A restriction to predicative methods has been a hallmark of positions *critical* of mathematical Platonism.

Thus, in the paper that introduced many of us to the idea that the great contending schools in twentieth-century philosophy of mathematics could be seen as positions on the question of universals, Quine describes a formulation of ramified type theory as representing a conceptualist viewpoint, far more restrained in its postulation of *abstracta* than that of the lotus-eating Platonists. Quine observes that the system is 'closely akin to Weyl's', and Weyl's position has also been characterized by Paul Bernays as *semiplatonism*, and as intermediate between the full Platonism of the classical set theoreticians and the intuitionism of Kronecker and Brouwer. The system described by Quine is also similar to that (derived from Wang) presented by Chihara as a nominalistically acceptable alternative to the Platonisms of Gödel and (the later) Quine.

All of these systems in effect take the natural numbers as individuals and accept mathematical induction as an additional form of inference beyond what is derivable from the ramified type-logic. (Quine does not make this explicit, but does say that the conceptualist 'tolerates elementary arithmetic and a good deal more', and compares his sample conceptualist system to others which do explicitly posit the arithmetic substructure at the bottom of the ramified hierarchy.) Even this marks a stronger assumption of abstract entities than what McMichael has shown to be needed for philosophy of science. The references to *abstracta* McMichael discusses can be provided by a predicative higher order *logic*, with no number theoretic assumptions incorporated. Such a logic is mathematically extremely weak. It can be proved consistent in a thoroughly finitistic fashion (Gentzen's finitistic proof of his *Hauptsatz* for first order logic extends to it in a straightforward way), no schema of infinity can be derived within it and, even when it is strengthened with an axiom of infinity, the principle of mathematical induction cannot be proved in it. If the far stronger systems of predicative mathematics are only conceptualist or semiplatonist, pure predicative higher order logic is as close to strict nominalism as one can come while still quantifying over abstract entities. Unlike the systems usually classed as Platonistic in the philosophy of mathematics, it postulates no structure that could be appealed to in deriving predictions or presenting explanations within physical science: a theory about physical objects formulated in a ramified higher order logic will be (by the *Hauptsatz*) a conservative extension of the corresponding first order theory. Nor, given the finitistic character of the consistency and conservative extension results, does it seem to present the same epistemological difficulties as more strongly Platonistic systems. (If all of mathematics were like ramified logic, Hilbert's formalistic program would have succeeded.) Since the abstract entities quantified over in a ramified system may all be thought of as propositions or properties definable in the language of physical science (though not necessarily of our *present* science), it does not involve us in the semantic mysteries about the possiblity of reference to abstract objects that have bothered critics of mathematical Platonism.

From the perspective of the philosophy of mathematics Field's version of physics is perhaps more objectionably 'Platonistic' than the simple quantification over *abstracta* commended by McMichael. For the supposedly concrete objects of Field's ontology include some (e.g. space-time points and arbitrary mereological sums of space-time points) which are not individually observable, even in principle. A sceptic might well claim that their postulation is not *physically* motivated, but is merely a trick to provide physical counterparts for the mathematical objects of classical analysis. But such a criticism of Field would be a topic for another paper.

University of Melbourne,
Parkville, Victoria 3052,
Australia

© ALLEN HAZEN 1985

REFERENCES

Field's book is *Science Without Numbers* (Princeton: Princeton U.P., 1980); the criticism suggested in the last paragraph is outlined in the reviews by Manders in the *Journal of Symbolic Logic* (1984) and Malament in the *Journal of Philosophy* (1982). Charles Parsons's observations about predicative higher order quantification is in his 'Sets and Classes', now reprinted in his *Mathematics in Philosophy* (Ithaca: Cornell U.P., 1983). The reference to Quine is, of course, to 'Logic and the Reification of Universals' in *From a Logical Point of View*. Wang's system is described in 'The Formalization of Mathematics', reprinted with several related papers in his *Logic, Computers, and Sets* (New York: Chelsea, 1970; earlier edition under title *A Survey of Mathematical Logic*), and Chihara's discussion of it is in *Ontology and the Vicious Circle Principle* (Ithaca: Cornell U.P., 1973). Weyl's theory is presented in *Das Kontinuum* (Leipzig: De Gruyter, 1918); there is an Italian translation, *Il Continuo* (Naples: Bibliopolis, 1977) but none in English. Bernays's evaluation is in his 'On Platonism in Mathematics', reprinted in Benacerraf and Putnam's *Philosophy of Mathematics* (Cambridge: Cambridge U.P., 1983): I have not been able to examine Bernays's original text, and the term 'semiplatonism' is the translator's.

REJOINDER TO TENNANT

By ALAN WEIR

IN an earlier article in this journal – 'Truth Conditions and Truth Values' (Weir [1]) – I put forward two criticisms of a form of anti-realism espoused by Neil Tennant in 'Is this a proof I see before me?' (Tennant [1]). Firstly I sketched a more radical anti-realism than Tennant's according to which one is only warranted in asserting that speakers understand S when one is warranted in asserting that they could effectively decide S's truth value. If one has grounds for asserting that speakers can discriminate correct from incorrect proofs of S, when they are presented with them, then this justifies one in denying that they do not understand S. But from an intuitionist point of view, of course, this justified denial entails that speakers understand S only if the property of understanding S is decidable, which it clearly is not. This more radical position involves a pleasant symmetry between determinacy of meaning and of truth value (symmetry between determinacy of meaning and of truth conditions is uninteresting where the latter is taken to be an ingredient of the former). I concluded that Tennant owes us an argument explaining why his type of anti-realism is more tenable than the more radical one (Weir [1] p. 178).

Tennant in his reply (Tennant [2] p. 104) declares that of course possession of proof-checking abilities entails grasp of meaning and that he is at a loss to understand why I should think otherwise. I am at a loss to understand why he should think that he has met my challenge, since he has produced no arguments, merely an example of proof-checking ability together with a reiteration of the disputed claim. But the existence and determinacy of proof-checking abilities, as well as proof-searching ones, which Tennant also mentions, is not in doubt; what is in doubt is whether a warrant for attributing such abilities suffices for attributing grasp of meaning rather than merely for refuting a denial of meaningfulness. Tennant provides no justification whatsoever for his confidence that the former rather than the latter is the case.

It is true that it is not clear quite what would settle this dispute, what the correct grounds are for asserting that specified conditions provide grounds for asserting that speakers understand a given sentence. I myself would reject both the radical position and Tennant's weaker verificationist answer. But I think that this rejection could only be made plausible in the context of a comprehensive account of the relations between meaning, truth conditions, belief and non-mentalistically described behaviour. Tennant seems to think that his rejection of the radical position is too obviously correct even to need justification. Intuition, however, seems a poor guide in such a complicated area.

Tennant may reply that he is under no obligation to provide positive grounds for his position, but merely to show that it is internally coherent. I have some sympathy with this reply since it is in keeping with the strategy of 'naturalized semantics' I sketched in [1]. By 'naturalized semantics' I meant the subproject of naturalized epistemology concerned with explanation of our knowledge of meanings, truth conditions, validity and the like. As with epistemology generally, it is foolish to cripple such an investigation by forbidding the use in it of the full range of our theories and rules of inference. Notwithstanding this, the task of showing, even using the full panoply of our logical laws, that the conception of truth and validity central to these laws is one appropriate to what we know of our epistemic capacities is far from trivial. Success in this task, then, would provide a non-trivial justification of fundamental logical laws.

Tennant, however, as a revolutionary challenging our prejudices in favour of classical logic, cannot be content merely with the internal coherence of his own position. And even if he undermines our confidence in classical logic by displaying a conflict between the concepts of truth and validity it employs and plausible constraints on our knowledge of meanings, the onus still lies with him to persuade us, disoriented and unsure of our intuitions, to join his sect rather than a rival band of radicals (of which, given the fissiparous nature of revolutionaries, there is no shortage). More fundamentally, Tennant needs to explain why a classicist's attempt at a justification, via natural semantics, of classical logic must fail. The bearing of my second criticism was that, at any rate in the restricted context of sentences which admit of conclusive verification, it would not. Tennant holds that a particular sentence is not neither true nor false. I apply double negation elimination to conclude that it is either true or false. 'What legitimate objection can be made' to this move, I asked ([1] p. 178). Significantly, Tennant offers none.

He does claim that the project of behaviouristic justification of realistic semantics I outlined, appealing to the internal structure of sentences and patterning of responses among complex sentences and their constituents, cannot handle quantification (Tennant [2] pp. 104-5). In [1] I argued merely that such a project cannot be written off *a priori* but that proper assessment of it would require much more extensive investigation than was possible in a brief paper. I had tried to do so in a paper subsequently accepted for publication (Weir [2] especially section 11) and have tried further in [3] (sections 5.1 and 5.3 in particular). Since Tennant challenges me on this point, I will conclude by sketching an approach to handling quantification classically, emphasizing that only a sketch is possible and the many stiff obstacles in the way of providing an adequate account require prolonged treatment.

As Tennant remarks, it is not possible to pattern responses to a generalization together with each of its constituents since there are (usually) infinitely many of these. But one can consider all the dispositions a speaker has to assent to the generalization together with a single substitution instance. *Pace* Kripke (p. 26), I find no impossibility in there being infinitely many of these (he is wrong in general in thinking that finite objects cannot have infinitely many dispositions: a charged particle has infinitely many dispositions of the form 'would accelerate at rate *r* under electro-magnetic force *F*' dependent on the size of the force). We can then require that speakers pattern their responses to these pairs of sentences appropriately. For the existential quantifier, one will require that the generalization always receives at least as great a degree of belief as the instance; for the universal, no greater a degree of belief. One might say that this behaviour is compatible with interpreting the quantifiers as sentential operators – 'maybe *F*' in the case of existential quantification, which never receives a less definite response than its immediate constituent, 'definitely *F*' for the universal quantifier, which never receives a more definite response.

But this is ruled out by the fact that the immediate constituent of the quantified sentence is not itself a closed sentence.

Still, no determinate function is expressed by operators obeying those constraints alone. But suppose that, in addition, we could train speakers so that, were they to be disposed to give decisive responses to each instance, then, whenever they responded more strongly to existential generalization than instance (or less strongly in the case of universal generalization) they would volunteer further instances to which they attached the same truth value as they did to the generalization. In such a case, it seems to me, the realist would be justified in interpreting the quantifiers as infinitary truth functions, maximizing, in the existential case, minimizing, in the universal case, over all the (possibly infinite number of) values of substitution instances.

If, further, we take the sentences to be queried to be not the generalization and each of its instances in turn but the generalization and a single instance with a parameter substituted for the quantified variable, we can bring into view objectual quantification. Instead of infinitely many dispositions towards pairs consisting of generalization and instances which differ syntactically, we consider infinitely many dispositions towards identical pairs of sentences but with the speaker understanding the parameter as referring to a distinct object in each case – one which, actually, may be forever unnamed and unthought of.

I suggest, then, the following criterion for grasp of universal quantification counting as a distinct instance each reading of F^x_a (the result of uniform substitution of all free occurrences of x in F by a term a not occurring therein) with the parameter 'a' assigned a different referent:

> S understands universal quantification iff for all predicates F in S's language, and for all objects x, if S were to understand a as referring to x, and $(x)Fx$ and F^x_a were to be queried, then S's utterances would be such that the degree of belief attached to the former will always be less than or equal to the degree of belief given to F^x_a. Moreover, were S disposed to make fully decisive responses to all instances of $(x)Fx$ then S would be able to produce an instance believed true to the same degree as the generalization, whenever the queried instance was not so believed.

The clause for existential quantification will be similar with 'greater than or equal to' substituted for 'less than or equal to'.

It goes without saying that there is a host of difficulties crowding round this account: can the notion of determinate linguistic dispositions be made sense of; or the notion of degrees of belief; can the account cope with the fact of error, from slips of the tongue to false belief; can the notion of decisive response be made behavioural sense of, indeed, can the notion of an intentional response to a sentence of any type be made sense of behaviourally? I try to answer these points in [3] (especially 3.3 and 4.2). Here all I would say is that they arise just as seriously for anti-realists whose rejection of realism is based on behaviouristic leanings in philosophy of language; yet nowhere, to my knowledge, is there a serious attempt by anti-realists to deal with them. Indeed these difficulties become intractable, it seems to me, if the background theory of one's naturalized semantics is anti-realist. How, for example, can one make sense of determinate dispositions, grounding determinate meanings, if subjunctives, being undecidable, lack stable truth values? Perhaps appeal to defeasible assertibility conditions will resolve this problem, but I doubt it.

The sketch of behaviourist realist semantics is not intended to persuade those who do not accept classical logic but to assuage the self-doubts of realists that the classical conception of quantifiers as expressing infinitistic truth functions, and hence capable of generating undecidable sentences, clashes with plausible views about how we could grasp such truth functions. (Of course if, as most philosophers do, one spurns a behaviouristic approach then present day anti-realist arguments become wholly unpersuasive.) I do not expect that the detail of my, or any, account will prove finally satisfactory. What one should look for in these areas, it seems to me, is a unified conception which yields specific solutions in a wide range of areas which are not obviously miles off target. I think this can be achieved in respect of the project of behaviouristic naturalized semantics for classical logic. At any rate, I have still not seen a counterargument of sufficient weight to justify abandonment of classical logic for intuitionist logic, with all the debilitating effect that would have on our reasoning, our mathematics and, through it, on our physical science.

Balliol College, Oxford OX1 3BJ

© ALAN WEIR 1985

REFERENCES

Saul Kripke, *Wittgenstein on Rules and Private Language*, Blackwell, Oxford 1982.
Neil Tennant [1] 'Is This a Proof I See Before Me?' *Analysis* 41.3, June 1981, pp. 115–19.
[2] 'Were Those Disproofs I Saw Before Me?' *Analysis* 44.3, July 1984, pp. 97–105.
Alan Weir [1] 'Truth Conditions and Truth Values' *Analysis* 43.4, October 1983, pp. 176–80.
[2] 'Realism and Behaviourism' *Dialectica* forthcoming.
[3] 'The Philosophy of Languages of W. V. Quine' Doctoral Thesis, Oxford, 1984.

BUCKNER QUOTING GOLDSTEIN AND DAVIDSON ON QUOTATION

By J. VAN BRAKEL

IN (Goldstein on quotation, ANALYSIS 44.4, October 1984, pp. 189–90), Buckner concludes that the demonstrative account of quotation is wrong because *it is fundamental to the demonstrative account that the 'quoted' sample is not read.* In a case like Buckner's sentence (4):

(A) *The word 'Kuznetzin' denotes the former Russian ambassador*

we DO read the 'Kuznetzin' and *without it the sentence does not have the required sense.* (Everything that is printed in italics I quote from Buckner's note just mentioned.) Hence, according to Buckner, the demonstrative account *must be wrong.*

This is a gross misrepresentation of the demonstrative account. Nowhere is it said, by either Goldstein or Davidson, that we do not read *the 'quoted' sample.* Buckner quotes Goldstein saying that *one can think of... quotation marks as arrows pointing to a sample which COULD be remote from the sentence*; and he quotes Davidson saying that *we COULD easily enough remove the quoted material from the heart of the sentence* (my emphasis in both cases). As Buckner shows himself (his sentences (2) and (3)), my sentence (A) (his (4)) can be easily rephrased, without changing the sense, so that the word

(B) Kuznetzin

is not read as part of the sentence. For example, instead of (A), I can say:

(C) the word (B) denotes the former Russian ambassador

However, whatever device we use to communicate something about the word (B), the sense of what is to be communicated is only obtained by the reader or listener if he has examined a token of (B). Buckner puts much emphasis on the fact that *first we read the quoted material; THEN we look back to examine it* (his emphasis). If I guess Buckner's intended meaning correctly, this would seem to be rather trivial: we can only start examining something after we have received the instruction or suggestion to do that and hence we have to read at least as far as the second quotation mark in (A), before we can start examining. If I go by the literal meaning of what Buckner says, it would seem that he misinterprets (A), because he says that *if we bracket off the quoted word* in (A), *then we must read*:

(D) *The word denotes the former Russian ambassador* *which does not make a great deal of sense.* But (D) is not the way we must read (A); neither on the demonstrative account, nor on any other sensible account. If we consider the quotation marks themselves as bracketing-off devices, the subject of (A) functions as an indexical description pointing to a part of itself. A similar situation arises in a phrase such as:

(E) the second word of this indexical description

where the reference of (E) is the second word of (E). On the other hand, if we take Davidson's suggestion that we can remove the quoted material from the sentence literally (which is what Buckner seems to suggest), then (A) does not reduce to (D), but to

(F) *The word '...' denotes the former Russian ambassador*

which on Buckner's own account is a perfectly sensible thing to read, because it is only after examining the quoted material that we might become suspicious with respect to the factual meaning of (F).

Of course Davidson does not mean that we can just cut the quoted material from the sentence and leave it at that. What is meant is that we *could easily remove the quoted material from the heart of the sentence* by rephrasing the sentence (for example (A) to (C)), perhaps even adding a sentence. And this we can always do. Quotation marks can always be eliminated by using some other conventions that serve the same goal. (But this is not the case the other way round. In the sentence

(G) (G) is not true

it is not possible to use quotation marks to name (G) and this simple fact has caused a whole stream of literature trying to make sense of Tarski's muddled presentation of the paradox of the liar.)

I conclude therefore that Buckner's criticism of the demonstrative account of quotation marks is misdirected and his positive conclusion that *whether or not we EXAMINE the quoted word, we must first of all read it*, is trivial and might suggest, incorrectly, that before examining it we would know what the sentence means.

Much of the discussion about various accounts of quotation marks suffers from the worn out fallacy that tokens of the same sign (for example quotation marks) should have the same meaning in every context where they occur. Quotation marks have been put to various uses: to mention an expression, to display an expression, to mention *and* use an expression, to indicate that an expression which is used has been copied from another source, to use an expression in a metaphorical way, and so on. Each use asks for its own analysis and not, *a priori*, for a reduction to one primary use. It would be instructive, just for a change, not to use quotation marks at all for, say, ten years, and get used to the different uses they have been put to by using different devices for each type of use, as I have done in this note. (No quotation marks occur in this note except in sentences which I have copied from Buckner's note.)

State University of Utrecht,
P. O. Box 80,000,
3508 TA, Utrecht, Netherlands

© J. VAN BRAKEL 1985

TARSKI A RELATIVIST?

By HARVEY SIEGEL

RICHARD C. Jennings ('Popper, Tarski and Relativism', ANALYSIS 43.3, June 1983, pp. 118–23 – all page references are to this paper) argues that Popper mistakenly appeals to Tarski's semantic conception of truth in order to defend a realist conception of science, and that Tarski's analysis actually supports or constitutes a form of relativism. In this brief note I will suggest that Jennings is right about the first point, but mistaken about the second.

Jennings correctly and fairly describes Popper's appeal to Tarski in his (Popper's) defence of the notion of objective or absolute truth. As Jennings points out, Popper takes Tarski's semantic conception to be an analysis of the correspondence theory of truth, which he (Popper) takes to explicate the relation between statements and extra-linguistic facts. As Jennings argues, however, the semantic conception offers no special access to such extra-linguistic facts. Rather, it 'implies no more than that whenever a sentence such as "snow is white" is asserted or rejected we must be ready to assert or reject the correlated sentence, "the sentence 'snow is white' is true"' (p. 121). What correspond on the semantic conception are sentences in the object-language and sentences in the meta-language, not sentences and extra-linguistic facts. The semantic conception does not, and is not intended by Tarski to, 'establish the conditions under which we are warranted in asserting any given (empirical) sentence' (p. 121). Thus Popper is mistaken in interpreting Tarski's semantic conception as explicating the relation between sentences and extra-linguistic facts. That conception, rather, is neutral as between alternative epistemological stances with respect to that relation: as Tarski says, '... we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naive realists, critical realists or idealists, empiricists or metaphysicians – whatever we were before. The semantic conception is completely neutral toward all these issues' (quoted by Jennings, pp. 121–2). Thus Popper is mistaken in taking Tarski's conception as supporting or justifying scientific realism. Here Jennings is right: the semantic conception will not secure Popper's realism.

But Jennings suggests that Tarski's conception actually constitutes a form of relativism, and supports only the sort of realism characterized by Putnam as 'internal realism', which of course is no realism at all. Here Jennings makes the same error in interpreting Tarski that he rightly accuses Popper of making. Jennings writes: 'We should see the world that our metalanguage tells us about as a world our object-language creates. In this way we can see the semantic conception of truth as a form of relativism' (p. 123). But Tarski's theory tells us no such thing. The semantic conception does not in any way support the view that the world is *created* by our object-language. It tells us nothing about the relation between sentences of the object-language and extra-linguistic facts. That is precisely the point that Jennings has been belabouring against Popper: Tarski's theory is *neutral* as between alternative 'epistemological attitudes' (p. 121) regarding the existence of facts and the relation between facts and sentences. The semantic conception is 'compatible with a relativistic view of ontology' (p. 119), but it is equally compatible with a non-relativistic, realist view. It is, as Tarski and Jennings both maintain, neutral with respect to these issues. But if it is neutral, it cannot also be taken to support the view that our object-language *creates* the world. Jennings is trying to have it both ways here: he is interpreting the semantic conception as being incapable of supporting Popper's realism, yet capable of supporting Jennings' own version of relativism. But he cannot have it both ways: the semantic conception cannot support Jennings' relativism for the very same reason that it cannot support Popper's realism: it is neutral. In particular, the semantic conception cannot legitimately be taken as showing that the world is created by our language, or that relativism is the correct 'epistemological attitude'.

Thus Jennings is right regarding Popper. Tarski's semantic conception does not imply realism, or force realism upon us. But neither does the semantic conception imply relativism, or force relativism upon us. It is, as Jennings sees in the one case but is blind to in the other, neutral.

University of Miami,
Coral Gables, Florida 33124,
U.S.A.

© HARVEY SIEGEL 1985

SAINSBURY ON DENYING A FREGEAN CONCLUSION

By STIG ALSTRUP RASMUSSEN

IN 'Rejoinder to Rasmussen' (ANALYSIS 44.3, June 1984: the rejoinder is to my note 'Sainsbury on a Fregean Argument' in the same issue), R. M. Sainsbury defends his contention that the argument adverted to in the title of his 'A Fregean Argument for the Distinctness of Sense and Reference' (ANALYSIS 43.1, January 1983) is inconclusive. I shall concentrate on the following version of the argument, using 'r(e)' and 's(e)' as abbreviations for 'the reference of e' and 'the sense of e' respectively:

(1) $r(a) = r(b)$.
(2R) X believes that $a = a$.
(3R) X does not believe that $a = b$.
(4R) If $s(e) = s(e')$, then ('X believes that $a = e$' is true) iff ('X believes that $a = e'$' is true).
(5R) $s(a) \neq s(b)$ (from 2R, 3R, 4R).

The Fregean conclusion follows that

(6R) Co-referentiality of names does not in general ensure sameness of sense.

Sainsbury now makes it clear that the Fregean's interlocutor is not to be depicted as entering open-mindedly into a discussion as to whether or not (5R) holds; his strategy is that of simply denying (5R) to see if this move will land him in absurdity. Sainsbury argues that it will at least not obviously do so. For, firstly, the interlocutor may deny (3R) on the strength of a distinction between the thought, or proposition, expressed by a sentence and what mastery of the use of the sentence involves a grasp of; and, secondly, he may join the Fregean in endorsing (3R), and take (2R) and the denial of (5R) to combine with (3R) to refute (4R).

Concerning the second strategy, Sainsbury's rejoinder gives me cause to add nothing to my point, in the previous paper, that (4R) must hold in virtue of the fact that our reason for introducing the notion of sense in the first place makes (4R) a constraint on that notion. Matters stand differently in respect of the first strategy. In his rejoinder, Sainsbury elaborates that strategy along these lines. True, we all have misgivings about ascribing to X, in cases such as the one involving 'Hesperus' and 'Phosphorus', a belief that $a = b$. However, it is open to us to regard (3R) as strictly speaking false, since its negation can without obvious absurdity be viewed as *misleading*, albeit literally true.

There may be no easy way of showing that this strategy commits the anti-Fregean to obvious absurdities. On the other hand, the strategy fails to make good that there is no need for a distinction akin to the Fregean one: as far as the analysis of linguistic mastery is concerned, the strategy merely serves to keep semantics simple at the cost of complicating what we may call pragmatics – in much the same way as does the Gricean treatment of seemingly non-truth – functional logical constraints (on this question, the reader may wish to consult R. C. S. Walker, 'Conversational Implicatures', in S. Blackburn ed., *Meaning, Reference, and Necessity*, Cambridge 1975). For the Fregean may, for the sake of argument, grant the proposed distinction between the proposition expressed by a sentence, as given by its (purely extensional) truth condition, and what is involved in linguistic mastery of the sentence. If 'sense' is used to signify what is expressed by linguistic expressions, the Fregean will also admit that the proposed strategy offers a way of denying (6R). However, he will go on to argue that this leaves us with the initial difficulty of explaining how the co-referentiality of 'a' and 'b' is compatible with the fact that we have misgivings about ascribing to X the belief that a = b, in cases where we confidently ascribe to X the belief that a = a. Given the Fregean's concessions to his interlocutor, the difference between 'a' and 'b' cannot now be alleged to be one of sense; but there must be *some* difference to do with the fact that one may master the use of both names without knowing that they refer to the same object. Suppose the Fregean accepted the terminology proposed by Sainsbury, even to the point of endorsing that the difference between 'X believes that a = a' and 'X believes that a = b' is that the latter is misleading, whereas the former is not. He could then go on to argue as follows, using 'u(e)' as short for what one has to know about e in order to know to what it refers:

(1) $r(a) = r(b)$.
(2S) 'X believes that a = a' is not misleading.
(3S) 'X believes that a = b' is misleading.
(4S) If $u(e) = u(e')$, then ('X believes that a = e' is misleading) iff ('X believes that a = e'' is misleading).
(5S) $u(a) \neq u(b)$.

From this follows the pseudo-Fregean conclusion that

(6S) Co-referentiality of names does not in general ensure sameness of what it involves to know to what the names refer.

There would now be a further question as to what exactly the notion abbreviated by 'u' is, i.e. how that concept dovetails with other notions in play in the analysis of language and thought. It is obvious from the outset that it cannot play all the roles envisaged by the Fregean for the concept of sense. In particular, the deliberate dissociation of the concept from that of truth ensures that it will have no central role to play in the analysis of propositions. However, if the argument (1) through (5S) is acceptable, there can be no denying the need for a notion other than a purely reference-theoretical one in terms of which linguistic mastery may be analysed.

The fact deserves emphasis that Fregeans will of course regard the terminology used in the above argument with suspicion: no Fregean would concede that the difference between 'X believes that a = a' and 'X believes that a = b' is merely one of misleadingness. However, the Fregean is not committing himself to make this concession, in the sense normally carried by the word 'misleading': he is merely adopting Sainsbury's terminology to express his own *Fregean* distinction between what is involved in mastery of 'a = a' and 'a = b' respectively. As long as this is the distinction captured, it does not matter that, by adopting Sainsbury's vocabulary, the Fregean may have conceded that what is involved in mastering an expression is a notion pertaining to pragmatics. By the same token, there can be no protesting against (4S), on the ground that sameness as to what is involved in mastery does not ensure preservation of misleadingness, in the ordinary sense: 'misleading' is used here in a very special sense. In fact, (4S) may legitimately be viewed as expressing a truism, for much the same reason which led to take (4R) to express a constraint on the notion of sense.[1]

University of Copenhagen,
Koebmagergade 50,
1150 Copenhagen K,
Denmark

© STIG ALSTRUP RASMUSSEN 1985

---

[1] I wish to thank the Editor for pointing out inadequacies in a previous version.

THE ONLY X AND Y PRINCIPLE

By H. W. NOONAN

A PLAUSIBLE principle governing identity is that whether a later individual, y, is identical with an earlier individual, x, cannot ever merely depend on whether there are, at the later time, any better candidates than y around for identity with x. I call this principle 'the only x and y principle'. In a previous paper ('Wiggins, Artefact Identity and "Best Candidate" Theories' ANALYSIS 45.1, January 1985, pp. 4–8) I argued that to reject it is to commit oneself to self-evident absurdities. In this paper my concern is to provide a reformulation of the only x and y principle which dispenses with the metaphor of 'best candidature', and makes it more obvious, I believe, why that principle is intuitively so compelling. This reformulation appeals to a variant of Geach's notion of a mere Cambridge change (Geach, *God and the Soul*, Routledge, 1969, p. 71).

In order to make the point of this reformulation clearer, however, it will be helpful to begin by considering an argument which uncovers some absurdities the 'best candidate' theorist is committed to in addition to those I drew attention to in my previous paper. The reformulation of the only x and y principle will then emerge in the course of explaining why a superficially plausible rejoinder by the 'best candidate' theorist to this argument is in fact ineffective.

As I elaborated in my previous paper, the 'best candidate' theorist's approach seems most plausible in a case like that of the Ship of Theseus, described by Hobbes. In thinking about this case there are three relevant situations to consider, which we can picture as follows:

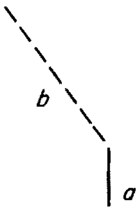

Figure 1

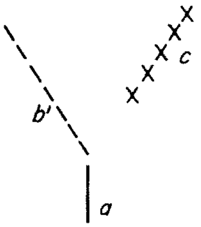

Figure 2

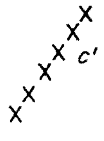

Figure 3

Here the second situation is the one described by Hobbes. The first situation is that in which the Ship of Theseus undergoes repair and replacement of parts, but the replaced parts are simply discarded or destroyed, and the third situation is that in which no repair or replacement work takes place, but the Ship of Theseus is dismantled plank by plank and later reconstructed. In each figure the continuous line represents the history of the Ship of Theseus before repair and replacement work begins. The line of dashes in Figures 1 and 2 represents the history of the continuously repaired ship, and the line of crosses in Figures 2 and 3 represents the history of the ship reconstituted from the original planks of the Ship of Theseus, i.e. the plank-hoarder's ship.

If we rigidly designate the ship originally referred to in all three situations as 'the Ship of Theseus' by 'a', the ship undergoing continuous repair in situation 1 by 'b', the ship undergoing continuous repair in situation 2 by 'b'', the ship reconstituted from the original planks in situation 2 by 'c', and the ship reconstituted from the original planks in situation 3 by 'c'', then, if we adopt the 'best candidate' approach to artefact identity we can, in conformity with the plausible view that both the continuously repaired ship and the plank-hoarder's ship are candidates for identity with the Ship of Theseus, though the former has the stronger claim, assert that in situation 1, $a = b$, in situation 2, $a = b'$ and $a \neq c$, and in situation 3 $a = c'$. We are then committed to saying that whether or not $c = c'$, the events which constitute the origin of c in situation 2, i.e. the reassembly of the hoarded planks into a ship, do not constitute the origin of that or any ship in situation 3, since – whether or not $c = c'$ – $c' = a$ which came into existence much earlier. This illustrates the first absurd consequence of the 'best candidate' approach which I wish to draw attention to: *events which constitute the origin of some entity of a certain kind in one situation, may not constitute the origin of that, or any entity of the kind, in a second situation even though all the events constituting the history of that entity in the first situation remain present in the second.*

Again, if we go along with the 'best candidate' approach and accept the statement given in the last paragraph of the identities and non-identities obtaining among a, b, b', c and c' in the Ship of Theseus case, we are committed to saying that two events in the history of ship c', i.e. ship a, in situation 3, one occurring before the disassembly and reconstruction of the ship, and one occurring after, will fail to be common parts of the history of that or any single ship in situation 2, even though both they, and all the events which were parts of the history of ship c' in situation 3, remain present in situation 2. This illustrates a second absurd consequence of the 'best candidate' approach: *two events may be parts of the history of a single entity in some situation, but may fail to be parts of the history of that, or any single entity of the kind, in a second situation in which both they, and all the events which were parts of the history of the entity in the first situation, remain present.*

At this point, however, a defender of the 'best candidate' approach may object that while these propositions are admittedly absurd, he is not committed to them. For in supposing that he is I have just *assumed* that one can decide whether an event which is part of the history of some entity in one situation is present in another without first deciding whether that entity itself is present in the second situation. But event identity is so tied up with object identity as to render this impossible.

However, I do not see that event identity *is* so tied up with object identity as to render this impossible. Not only are the events in the history of the plank-hoarder's ship in situation 2 qualitatively identical with the events in the history of the plank-hoarder's ship in situation 3, the plank-hoarder's ship in situation 2 is built of exactly the same planks, manned by exactly the same crew and sails over exactly the same seas as the plank-hoarder's ship in situation 3. Moreover, the 'additional' events in situation 2, i.e. those which do not have even qualitative counterparts in situation 3, have no causal effect whatsoever on the events in the history of the plank-hoarder's ship in that situation (this is part of the description of the situation). Given all this, it seems evident that while there may be *a* notion of event identity in accordance with which to determine whether the events in the history of the plank-hoarder's ship in situation 3 are identical with the events in the history of the plank-hoarder's ship in situation 2 one must first decide whether the plank-hoarder's ship in situation 3 is identical with the plank-hoarder's ship in situation 2, this is certainly not true of our ordinary notion of event identity.

At this point it is helpful to introduce Geach's distinction between real and mere Cambridge changes. An object undergoes a Cambridge change if a proposition about it changes in truth-value. Thus Theaetetus underwent a Cambridge change when he grew taller than Socrates, and so did Socrates. But the change in Theaetetus was a real one, whereas the change in Socrates was a *mere* Cambridge change. Mere Cambridge changes are Cambridge changes which are not real ones, and though the distinction between real and non-real changes is hard to define, our grasp of it, as manifested in our non-collusive agreement on examples, seems sufficiently firm to make the notion of a mere Cambridge change useful in philosophical discussion.[1] Now given the notion of a mere Cambridge *change* one can evidently define a family of related notions. Thus a mere Cambridge *property* is one that can be gained or lost by a mere Cambridge change. Two objects are *similar* in a mere Cambridge way if they share a mere Cambridge property. Two objects *differ* in a mere Cambridge way if one possesses a mere Cambridge property the other lacks. Finally we can define the notion of *two possible situations differing in a mere Cambridge way with respect to what happens in a certain location*. This will be so in the case of two situations S and S′ with respect to what happens at a location L just in case the only changes occurring at L in either situation not occurring there in the other are mere Cambridge changes, and the only properties instantiated at L in either situation not instantiated there in the other are mere Cambridge properties. And now I can put the point I was making at the end of the previous paragraph about our ordinary notion of event identity in a different way: with respect to what happens at the location of the plank-hoarder's ship in situations 2 and 3 the difference between the two situations is a mere Cambridge difference, and our ordinary notion of event identity is such that it is a sufficient condition of the events occurring at a certain location being identical in two situations that, with respect to what happens at that location, there is a mere Cambridge difference between the two situations. But if this is our ordinary notion of event identity it follows that the 'best candidate' theorist's envisaged rejoinder to my argument fails.

Finally I can offer the non-metaphorical reformulation of the only x and y principle I promised at the outset of the paper, a reformulation which, for me at least, reveals much more clearly than its original formulation the intuitive force of that principle. Let us call the sufficient condition of trans-world event identity stated at the end of the previous paragraph the 'Cambridge criterion'. Then the only x and y principle may be reformulated as follows: if two events are parts of the history of a single entity of a kind in one situation then they must also be parts of the history of a single entity of the kind in any second situation in which, as judged by the Cambridge criterion, both they, and all the events which are parts of the history of the entity in the first situation, remain present.

---

[1] But perhaps, as suggested by Shoemaker 'Causality and Properties' (in P. Inwagen (ed.) *Time and Cause*, D. Reidel, 1980), it can be defined in an epistemological way. To verify in the most direct way that an object has undergone a *real* change at a certain place and time one must make observations and tests in the vicinity of that place and time, whereas to verify in the most direct way that an object has undergone a mere Cambridge change at a certain place and time it will be necessary to make observations and tests remote from that place and time, and observations and tests made at that place and time will either be irrelevant or insufficient.

So formulated I find the only x and y principle undeniable. Nevertheless acceptance of it has far reaching consequences. But the exploration of these must be left to another paper.[2]

University of Birmingham,
P.O. Box 363, Birmingham B15 2TT

© H. W. NOONAN 1985

---

[2] For such an exploration, in the context of a discussion of Robert Nozick's 'Closest Continuer' theory of identity, see my 'The Closest Continuer Theory of Identity', forthcoming in *Inquiry* 1985.

SOME RADICAL CONSEQUENCES OF GEACH'S LOGICAL THEORIES

By JAMES CAIN

PETER Geach's views of relative identity, together with his account of proper names and quantifiers,[1] while presenting what I believe is an inwardly coherent and consistent account, presents us with radical consequences regarding what arguments are to be accepted as valid. For example, consider the argument:

(1) All men are mortal
Fido is a man
Thus, Fido is mortal

in which 'Fido' names a dog. While this would normally be thought to be a valid argument with a false premise, we shall see that on Geach's theories it turns out to be invalid. In fact for every form of general categorical syllogism we can produce an argument of that form which on Geach's theories turns out invalid.

---

[1] I will only be concerned with the theory of restricted quantifiers worked out in the main body of *Reference and Generality*, third edition (Ithaca: Cornell University Press, 1980), not with either his theory of unrestricted quantifiers or the method of interpreting restricted quantifiers mentioned in the appendix to *Reference and Generality.*

Let us see how this situation arises. We first look briefly at Geach's semantics. According to Geach every proper name is associated with a criterion of identity which could be expressed using a substantival term; e.g., 'Fido' is, let's say, associated with the criterion of identity given by 'same dog': this can be expressed by saying that 'Fido' is a name *for* a dog (we may go on to say that it is a name *of* a dog if it is non-empty; *Reference and Generality* p. 70). Restricted quantification is in turn explained in terms of proper names. Geach says:

> '*F*(some *A*)' is true iff '*F*(*a*)' is true for some interpretation of '*a*' as a name *of* and *for* an *A*;
>
> '*F*(any *A*)' is true iff '*F*(*a*)' is true for any interpretation of '*a*' as a name *of* and *for* an *A*.
>
> (These schematic conditions are of course to be applied to actual concrete examples by replacing the letter '*A*' with some actual general term both within and outside the quotes.) (*Reference and Generality* p. 206)

Since we will eventually also consider phrases of the form 'No *A*', let us supplement Geach's definition for restricted quantifiers with:

> '*F*(no *A*)' is true iff '*F*(*a*)' is true for no interpretation of '*a*' as a name *of* and *for* an *A*.

Now we turn to Geach's views on identity. According to Geach's views we may have cases in which *a* and *b* are the same *A* yet we may be able to say conflicting things about them. One example involves heralds; Geach introduces us to heralds as follows:

> In England the members of the Heralds' College have a range of special names, such as 'Garter' 'Rouge Dragon', 'Portcullis'. These are grammatically, and I think logically, proper names. Each of these is a name *for* a herald; by repeated use of it we have in mind not one and the same man, but one and the same herald – the filler (whoever he may be) of one and the same heraldic office. ('Existential or Particular Quantifier' in Weingartner and Morscher (eds.) *Ontology and Logic*, Dunker and Humblot, 1979, p. 146)

In the typical case in which just one man holds a given office in the Heralds' College we will be able to say, according to Geach, that the herald is a man. However, at different times one herald may be different men, as in the following example of Geach:

> At a given point of time the man Hilary Handel is a herald, in fact he is (the same herald as) the herald Sable Basilisk; conversely, at the same point of time the herald Sable Basilisk is a man, in fact he is (the same man as) the man Hilary Handel . . . one and the same herald, say Sable Basilisk, may at different times be (be the same man as) different men. (*Reference and Generality* p. 217)

Even though, let's suppose, Sable Basilisk and Hilary Handel are now the same man, conflicting predicables might apply to them, e.g., '____ has existed for more than fifty years' might apply to Sable Basilisk, but not to Hilary Handel.

Now consider the argument:

(2) All men are mortal.
Sable Basilisk is a man.
Thus, Sable Basilisk is mortal.[2]

In the light of the fact that heraldic posts may be held by a succession of men, and therefore a herald might exist forever (as far as logic is concerned), it is clear that the conclusion does not follow from the premises. The argument turns out to be invalid. (At this point we need to see that the first premise comes out true, according to Geach's semantics, as long as '*a* is mortal' is true for any possible reading of '*a*' as a name *of* and *for* a man: Sable Basilisk might exist forever and the first premise still hold since 'Sable Basilisk' is not a name *for* a man.) We now also see why argument (1) must also be treated as invalid according to Geach's theory: it is logically on a par with argument (2).

This sort of consideration also shows the following arguments to be invalid (imagine the case in which no herald ever ceases to exist, yet all heralds are men):

All men are mortal.
All heralds are men.
Thus, all heralds are mortal.

All men are mortal.
Some herald is a man.
Thus, some herald is mortal.

And these would typically be taken to be valid syllogisms. Using our imagined example in which heralds never cease to exist, we could find arguments in which the premises come out true and the conclusion false corresponding to each form of general categorical syllogism in the first three figures traditionally regarded as valid. These arguments are merely slight variations upon our last two examples, so I need not list them. (In each case either 'mortal' or 'immortal' corresponds to the term which does not ever appear in subject position. 'Man' and 'herald' correspond to the other terms.)

Even Geach, up through his third edition of *Reference and Generality*, seems to have been unaware of the shake-up that his theories cause in what arguments we must recognize as valid. To give just one example, Geach takes the following to be a valid argument form: 'Socrates in an *A*; any *A* is a donkey; ergo Socrates is a donkey'. (*Reference and Generality* p. 143) Surely the validity of this form stands or falls with the validity of:

---

[2] If one feels uncomfortable using the predicable 'is mortal' when speaking of heralds, one may substitute the predicable 'will have ceased to exist'; or we may stipulate that in this paper 'is mortal' shall mean 'will have ceased to exist'.

a is an *A*.
All *A* are *F*.
Thus, *a* is an *F*.

But we have already shown arguments (1) and (2) to be invalid in Geach's system, and they are instances of the above with the order of the premises switched.

Consider further Geach's views on identity. Geach gives the following example which would confirm Wiggins' thesis *R*:

> A parliament intent on economy might decree that in the national emergency Norroy shall be the same man as Clarencieux; in that case 'Sir Hilary Handel', a name *for* a man, might be a name shared by two different heralds, Norroy and Clarencieux, and yet be a proper name of a man, since it names but one man and *ex hypothesi* each of these heralds will be the same man as Sir Hilary Handel. ('Existential or Particular Quantifier' p. 148)

Note that we could not, *in this case*, say that Sir Hilary was a herald. For it would be absurd to say that he was the same herald as just one of the two. Nor would he be the same herald as both, or by the transitivity and symmetry of 'is the same herald as' Clarencieux and Norroy end up being the same herald – but the example stipulates that they are not. The only alternative left if Handel is to be a herald is that he be a different herald from Clarencieux and Norroy, but this gives the unwanted result that we have three heralds. We must conclude that Handel is not a herald.

Let us extend this last example by imagining that the human race is all but destroyed, so that Handel is the only human left, and Norroy and Clarencieux are the only heralds left. Since Handel is not a herald we have:

No men are heralds

coming out true and yet its converse:

No heralds are men

coming out false since both Clarencieux and Norroy are men. Yet this pair of propositions would normally be regarded as an instance of an equivalent pair of schemata:

No *A* are *B*

and,

No *B* are *A*.

Consider also the argument:

No humans are heralds.
Some heralds are human.
Thus, some humans are not human.

Our example would make the premises come out true and the conclusion false, though this would typically be regarded as a valid syllogism of the fourth figure, being schematized:

No *A* are *B*.
Some *B* are *C*.
Thus, some *C* are not *A*.

In fact, using our example, we may provide invalid arguments exemplifying each form of argument generally recognized as valid in the fourth figure. So there are counterexamples to each form of general categorical syllogism generally recognized as valid.

This last example, by the way, provides a counterexample to the following, which Geach offers as a valid schema ('Names and Identity' in Guttenplan (ed.) *Mind and Language*, Oxford 1975, p. 140):

Every *A* is a *B* or is a *C*.
Every *B* is *P*; every *C* is *P*.
Thus, every *A* is *P*.

Consider this interpretation of the schema:

Every herald is a man or is a man.
Every man is not a herald; every man is not a herald.
Thus, every herald is not a herald.

Our example makes the premises true and the conclusion false.

Geach's theory not only drives us to treat as invalid arguments that would normally be treated as valid, but also to treat as valid arguments that would normally be treated as invalid. Consider the following argument in which 'Socrates' functions as a name *of* and *for* a man:

All men are mortal.
Thus, Socrates is mortal.

This argument is usually regarded as an enthymeme. On Geach's theory it should be regarded as strictly valid. Since, on Geach's view, it is part of the sense of 'Socrates' that it is a name *for* a man (*Reference and Generality* pp. 70–1), we are justified in concluding that Socrates is mortal, if we are given that all men are mortal. This is shown as follows: the premise comes out true iff '*a* is mortal' is true under any possible reading of '*a*' as a name *of* and *for* a man; 'Socrates' presents us with such a reading of '*a*' (Geach holds that empty names are inadmissible in logic – so we need not worry about 'Socrates' not being a name *of* a man); thus with 'Socrates' replacing '*a*' we get a truth whenever the premise is true. So the conclusion follows from the premise and the argument is valid.

Geach does provide an example in which we get a valid argument where it would normally be thought that we have an enthymeme (*Reference and Generality* p. 109). He allows that from the premise 'no man has admired any pig' we may infer 'the man *a* has not admired any pig', where, Geach says, he uses ' "the man *a*" as proxy for a term that is a name *of* and *for* a man'. In that case, since 'Socrates' is a name *of* and *for* a man, 'Socrates has not admired any pig' validly follows from the premise by itself; there is no need to add a premise stating that Socrates is or was a man.

The examples in this paper have not shown an inconsistency in Geach's system. That they are surprising examples I think shows us that we do not very readily know our way about logically, given Geach's style of restricted quantification together with his theory of relative identity.

Here I must mention one problem that has not been one of my concerns in this paper; it involves the use of schematic letters in displaying syllogistic patterns. Consider the sentence 'all men are mortal'. We have treated this sentence as having the form 'all *A* are *B*' where '*A*' and '*B*' are in the same syntactic category. Certainly this schematization is problematic ('all mortal are men' does not even make good sense).[3] But, in any case, we can recognize a common logical form in, say, the propositions: 'all men are mortal', 'all heralds are British', 'all dogs are plants'. How we adequately schematize this form will depend in part upon whether Geach's theories or some other logical theories are correct. But we do need some way to indicate what form of proposition or argument we are speaking about. To this end we have used the traditional schematizations because anyone should then know what form of proposition or argument we have in mind.

I believe that the general view among philosophers is that – however one might have to tidy things up to get a theory of syllogisms in proper formal shape – any argument is valid if we would now (prior to tidying things up) be willing to informally treat it as instantiating a traditionally accepted syllogistic pattern (with the exception of cases where validity depends upon universal quantifiers having been given existential commitment). We now see that if we accept Geach's claims regarding relative identity and restricted quantifiers, we must reject this general view. With the examples in this paper we can see how radically Geach's theories would change what we count as a valid argument.

University of Pennsylvania

© JAMES CAIN 1985

---

[3] Note that in my examples the only terms I have used in subject position are substantival terms, and counterexamples for each syllogistic form can be provided with the subject terms being only substantival terms.

LEMMON ON LOGICAL RELATIONS

By RALPH KENNEDY

E. J. LEMMON states, in his *Beginning Logic* (Nelson 1965), that that part of traditional logic concerned with the logical relations between propositions (equivalence, contradiction, contrariety, subcontrariety, implication and its converse) 'can be viewed as an (inadequate) attempt to list all possible relations between propositions which can be defined in terms of truth-values' (p. 70). In this paper I describe what seems to me a reasonable way of understanding the aim of the traditional account, given which it is by no means obvious that that account is in any way inadequate or incomplete.

I

Lemmon characterizes the relations he calls 'definable in terms of truth-values' by reference to the sixteen binary truth-functions (see Figure 1). Consider implication, for instance. 'A proposition A *implies* a proposition B if whenever A is true B is true (but not necessarily conversely)' (p. 69). Lemmon's (partial) characterization of this relation by reference to function (e) runs: 'If A and B are expressible in propositional calculus notation, the claim that A implies B is just the claim that the function of A and B defined by (e), namely $A \rightarrow B$, is tautologous' (p. 71). To cite one more example: '. . . to say that A and B are equivalent is to say that the function defined by (g), $A \leftrightarrow B$, is tautologous' (p. 71). Fourteen further relations are defined similarly by reference to the fourteen remaining binary truth-functions.

Before proceeding to the main business of this paper, we should note that Lemmon's way of defining these relations is unsatisfactory. Lemmon evidently wants 'A' and 'B' to range over what he calls 'propositions', not wffs of the propositional calculus. But if A and B are not wffs, then neither is $\ulcorner A \rightarrow B \urcorner$,[1] and it won't be clear what calling $\ulcorner A \rightarrow B \urcorner$ *tautologous* would mean. (According to Lemmon's definition, a tautology is a wff of the propositional

AB	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p
TT	T	T	T	T	T	T	T	T	F	F	F	F	F	F	F	F
TF	T	T	T	T	F	F	F	F	T	T	T	T	F	F	F	F
FT	T	T	F	F	T	T	F	F	T	T	F	F	T	T	F	F
FF	T	F	T	F	T	F	T	F	T	F	T	F	T	F	T	F

Figure 1

---

[1] The corner-quotes here indicate my assumption that '$A \rightarrow B$' must have been intended to denote the conditional with antecedent A and consequent B, i.e. $\ulcorner A \rightarrow B \urcorner$.

calculus which 'takes the value T for all possible assignments of truth-values to its variables' p. 68.)

We could get around this particular difficulty by talking not of tautologies but of their substitution-instances, stating, e.g., the condition for implication to be that $\ulcorner A \rightarrow B \urcorner$ be a substitution-instance of a tautology. For equivalence the condition would be that $\ulcorner A \leftrightarrow B \urcorner$ be a substitution-instance of a tautology, and so on for the remaining relations. But there would still remain an obstacle in the way of following Lemmon in taking 'A' and 'B' to range over *propositions*, conceived of as what sentences *express* (p. 6).[2] Expressions like '$\ulcorner A \rightarrow B \urcorner$' and '$\ulcorner A \leftrightarrow B \urcorner$' resist interpretation if 'A' and 'B' are taken to range over propositions. Propositions presumably do not have syntactic structure; hence, 'the proposition with antecedent A and consequent B' does not make sense. Perhaps the thing to do is to re-write the conditions along the lines of the following models: [implication] there are sentences A^* and B^* that express A and B, respectively, such that $\ulcorner A^* \rightarrow B^* \urcorner$ is a substitution-instance of a tautology; [equivalence] there are sentences A^* and B^* that express A and B, respectively, such that $\ulcorner A^* \leftrightarrow B^* \urcorner$ is a substitution-instance of a tautology; etc.

A more serious difficulty with Lemmon's definitions concerns his use of the restrictive phrase 'expressible in propositional calculus notation'. The role played by this restriction *seems* at first clear. One would not want to deny the claim, for instance, that 'Socrates is snub-nosed' implies 'someone is snub-nosed', even though 'Socrates is snub-nosed → someone is snub-nosed' is not a substitution-instance of a tautology. And this claim need not be denied if 'Socrates is snub-nosed' and 'Someone is snub-nosed' are deemed not 'expressible in propositional calculus notation'. But then, presumably, 'Socrates is snub-nosed and someone is snub-nosed' would not count as so expressible either (it's more complex than either of its constituents); hence, one should not be able to appeal to Lemmon's definition to show that 'Socrates is snub-nosed and someone is snub-nosed' implies, e.g., 'Someone is snub-nosed', despite the fact that the two are substitution-instances, respectively, of the wffs 'P & Q' and 'Q', the first of which truth-functionally implies the second.

There is no easy solution to this problem so far as I know, but we can sidestep it by abandoning the restriction and regarding our conditions (for implication, equivalence, etc.) as merely sufficient, rather than necessary and sufficient. We would then have, e.g.:

---

[2] Here I mean to discuss only obstacles specific to Lemmon's project. I shall not discuss the many general objections that could be raised against employing the conceptual machinery of propositions and proposition-expressing sentences. Perhaps this is a good place to note that, if we were not trying to stay as close as possible to Lemmon's formulations, we *could* think of 'A' and 'B' as ranging over 'eternal' sentences of a 'regimented' version of English; if we did so, we could leave our conditions as they now stand, avoiding the complications about to be introduced in the text.

A implies B *if* there are sentences A* and B* expressing A and B respectively such that $\ulcorner A^* \rightarrow B^* \urcorner$ is a substitution-instance of a tautology; etc. (It would be fairly easy at this point to move much closer to full definitions simply by replacing 'tautology' in these formulations by 'valid schema', the implicit reference being to a richer system of logic than the propositional calculus. But my present criticisms do not require this move.)

II

In calling the traditional account 'inadequate' Lemmon must have meant that it is arbitrary to focus one's attention on just the usual six relations, that there are no significant differences between this group and the remaining ten relations he mentions. But this is not correct. The six traditionally recognized relations do share a very significant property absent from the remaining ten, a property that it seems quite natural to connect with the notion of a 'logical relation': they and only they (of the sixteen relations enumerated by Lemmon) are capable of supporting certain key inferences involving their relata. I say that such relations have 'immediate deductive utility'. Let me spell out what is involved in a relation's having this property.

First, a relation R has immediate deductive utility only if: (1) for all A and B, if A bears R to B then an ideal logician who knows the truth-value of A will be able to come to know, by deduction, the truth-value of B, or vice-versa. This rules out the universal relation, that corresponding to function (a), since there exist pairs of propositions such that an ideal logician knowing the truth-value of either one could not deduce that of the other. Thus, at most fifteen of Lemmon's relations have immediate deductive utility.

If we required nothing more than condition (1) then we would have to say that each of the remaining fifteen relations had immediate deductive utility. But it is reasonable to require more. Suppose, for instance, that R is that relation holding between A and B if and only if B is expressible by a substitution-instance of a tautology. R then satisfies (1). But there is not much to be said for R's being a relation with immediate deductive utility. For one thing, R's status as a *relation* of any sort is marginal: the conditions necessary and sufficient for R's holding between two items are statable entirely in terms of 'internal' properties of just one of those items. A related peculiarity of R is that its holding between two propositions A and B *never* makes it possible for an ideal logician to come to know the truth-value of A *on the basis of knowing that of B* or vice-versa. Surely it makes sense to formulate our criterion of immediate deductive utility in such a way as to exclude R and other relations of its ilk. We can achieve this effect by adding as a second necessary condition: (2) for some A and B, A bears R to B but not even an ideal logician could deduce the truth-value of A without knowing that of B, nor could she deduce the truth-value of B without knowing that of A.

Of Lemmon's relations, only the six he calls 'traditional' would appear to satisfy both conditions (1) and (2).

We have seen that the universal relation does not satisfy (1). The null relation – that corresponding to function (p) – does not satisfy (2).

The six 'traditional' relations correspond to functions (i) (contrariety), (b) (subcontrariety), (g) (equivalence), (j) (contradiction), (e) (implication), and (c) (the converse of implication). It is clear that these all satisfy both conditions (1) and (2).

It remains to be argued only that none of the remaining relations satisfies (2). Some care is needed here, given that we (and Lemmon) have only *partially* characterized these relations. One cannot argue on the basis merely of *sufficient* conditions for a relation's holding between two propositions that there could *not* be two propositions so related satisfying a certain condition. But we *can* say this: if R is any one of the remaining relations (corresponding to function (d), (f), (h), (k), (l), (m), (n), or (o)) and propositions A and B are counted by our partial characterization of R as being related by R, then an ideal logician *could* deduce the truth-value of A without knowing that of B or vice-versa. In other words, given only our (and Lemmon's) partial characterization of R, there is no reason to affirm (2) of R. This is because, if ARB holds in virtue of that characterization, then at least one of A, B, $\ulcorner -A \urcorner$, or $\ulcorner -B \urcorner$ is already expressible by a substitution-instance of a tautology. For instance, if R is the relation corresponding to function (d), and ARB holds in virtue of our partial characterization of R, then A must be expressible by a substitution-instance of a tautology.

If, then, it is reasonable to view the traditional account as aiming to provide a list of relations useful in the sense defined by the conjunction of conditions (1) and (2), Lemmon has not succeeded in showing that account to be inadequate.

III

Two final remarks:

(1) Lemmon notes that according to the traditional account, 'if none of the six relationships holds between A and B then A and B are *independent*' (p. 70). This conception of independence as the absence of logical relation is so intuitive that it would seem worth trying to respect it in giving an account of logical relation. One could even view it as giving us a way of recognizing at what point our list was *complete*. Let us say that what it is for A and B to be logically *independent* is for each of the following to be logically possible: A and B are both true; A and B are both false; A is true and B is false; A is false and B is true. A list of logical relations could then be called *complete* just in case for any two propositions

A and B, A and B are independent if and only if A and B do not bear to one another any relation on the list. The traditional list of six relations appears to be complete in this sense, even redundant: our criterion would still be satisfied even if equivalence and contradiction were dropped. It cannot be said therefore that Lemmon's list is superior to the traditional one on the score of completeness.

(2) There are propositions which one might well want to call logically related but whose relation is not *specifically* recognized either by Lemmon's account or the traditional one. Consider, for instance, any propositions expressed by substitution-instances of 'P v Q' and 'Q v R'. Lemmon's account would say that these are related, but only by the relation *any* two propositions bear to one another. So again, whatever we want to say about such propositions, they do not point to a respect in which Lemmon's account is superior to the traditional one.[3]

Wake Forest University,
Winston-Salem,
North Carolina 27109, U.S.A.

© Ralph Kennedy 1985

---

[3] I am indebted to Professor Peter Geach for his helpful comments on a remote ancestor of this paper and to Professor Christopher Hill and the Editor of *Analysis* for much useful criticism of more recent drafts.

'IF A AND B, THEN A'

By E. J. Lowe

THERE is an almost universal assumption, amongst those currently writing on the logic of conditionals, that 'If A and B, then A' must be regarded as an axiom-schema (or at least theorem-schema) of any acceptable conditional logic. The point is of no small consequence, since it is often appealed to in arguing against the *transitivity* of conditionals, that is, against the validity of the inference-schema

(T) $A \to B, B \to C \vdash A \to C$

The argument turns on the fact that

(T*) $(A \& B) \to A, A \to C \vdash (A \& B) \to C$

is a substitution-instance of (T), so that if we additionally have, following the aforementioned assumption, the axiom-schema

(C) $\vdash (A \& B) \to A$

we may immediately conclude that the inference-schema commonly known as 'strengthening the antecedent'

(S) $A \rightarrow C \vdash (A \,\&\, B) \rightarrow C$

is valid. (See, e.g., Robert C. Stalnaker, 'A Theory of Conditionals', in E. Sosa (ed.), *Causation and Conditionals*, Oxford U.P., 1975, p. 173 and Geoffrey Hunter, 'Conditionals, Indicative and Subjunctive', in J. Dancy (ed.), *Papers on Language and Logic*, University of Keele, 1979, p. 116.) It is an embarrassment for any theory of conditionals to have to concede the validity of (S) because there are so many *prima facie* counterexamples to it, e.g. 'If Smith strikes that match, it will light; so if Smith strikes that match and at the same time holds it under water, it will light'.

But why should we accept the assumption? It won't do simply to point out the obvious fact that it is logically impossible for 'A & B' to be true and yet 'A' to be false, for this only goes to show that so-called *strict* conditionals of the form '$\Box((A \,\&\, B) \supset A)$' are logically true: and there are good and well-known reasons for denying that 'If . . . , then . . . ' sentences in natural language should be analysed as strict conditionals (much less, of course, as material conditionals). Nor will it do merely to observe that conditionals of the form 'If A and B, then A' very commonly appear, even to untutored intuition, just trivially true – e.g. 'If I go to the shop and buy a packet of tea, I shall go to the shop'. For in assessing the validity of the schema '$(A \,\&\, B) \rightarrow A$' it is not enough just to point to a number of plausibly true instances of it without regard to the question of whether these instances share some special feature in virtue of which they collectively give support only to some weaker claim. If '$(A \,\&\, B) \rightarrow A$' is a valid schema, *all* instances of it must of course be true, including (for example) instances in which 'A' and 'B' are *mutually incompatible* (e.g. where 'B' = '$\sim$A'). But in view of the fact that we rarely if ever encounter in ordinary discourse conditionals with (manifestly) inconsistent antecedents, appeal to 'linguistic intuition' in this area can have very little, if any, force. Thus, faced with the question of whether, say, the conditional 'If I go to the shop and I do not go to the shop, I shall go to the shop' is true according to untutored intuition, *we just have no answer*, for the simple reason that we never have occasion in ordinary discourse to consider the truth-value of such a conditional. At most, I suggest, the evidence of 'linguistic intuition' supports the validity not of the schema '$(A \,\&\, B) \rightarrow A$', but only of the inference-schema

(P) $\Diamond(A \,\&\, B) \vdash (A \,\&\, B) \rightarrow A$

Now, however, the additional endorsement of (T) cannot be shown to commit us to the objectionable (S), but at most only to the weaker principle

(W) $\Diamond(A \,\&\, B), A \rightarrow C \vdash (A \,\&\, B) \rightarrow C$

Here some may be inclined to retort that (W) is in fact no less embarrassing a commitment than (S) itself. For instance, it may be urged that someone asserting 'If Smith strikes that match, it will light' must still infer 'If Smith strikes that match and at the same time holds it under water, it will light', because such a person must surely concede that it is *possible* for Smith both to strike the match and at the same time to hold it under water. However, everything depends here on what sense of 'possible' is at issue. Obviously it is *logically* (and indeed *physically*) possible both to strike a match and at the same time to hold it under water. But notice that if someone *were* to assert 'If Smith strikes that match, it will light' we would surely expect him to have reason to suppose that Smith *would not* in fact both strike the match and at the same time hold it under water. After all, if one knew, say, that Smith had a quirky habit of striking matches under water one would presumably *not* think it reasonable to assert 'If Smith strikes that match, it will light'. (The assertibility of conditionals in ordinary discourse is notoriously context-dependent, in the sense that whether or not it is reasonable for a speaker to assert a given conditional in given circumstances commonly depends very much on what the speaker knows about those circumstances.) Now what I am suggesting is that if we consider *epistemic* (rather than merely logical) possibility, it is plausible to hold that someone asserting 'If Smith strikes that match, it will light' will indeed be ruling out the 'possibility' that Smith should both strike the match and at the same time hold it under water, and that *this* is why such a person does not commit himself to the truth of the conditional 'If Smith strikes that match and at the same time holds it under water, it will light': he is not committed because, though (W) *is* (I suggest) a valid inference-schema, he is not committed to one of the premises which (W) requires for the undesirable conclusion to follow.

An objection that might be raised at this point is that I have no right to restrict the sense of the modal operator in (W) to 'epistemic' possibility because I have no right to do so in (P). Surely, it may be said, even someone who has reason to suppose that Smith will not in fact both strike the match and at the same time hold it under water (i.e. who rules out the epistemic possibility of this conjunctive state of affairs obtaining), must still accept the truth of the conditional 'If Smith strikes that match and at the same time holds it under water, Smith will strike that match'. Moreover, given the validity of (T), such a person, if he also asserts 'If Smith strikes that match, it will light', is still committed to the undesirable conclusion 'If Smith strikes that match and at the same time holds it under water, it will light'; this, in fact, is just an application of (T*). (Compare John L. Pollock, *Subjunctive Reasoning*, Dordrecht: Reidel, 1976, p. 40 and Wayne A. Davis, 'Indicative and Subjunctive Conditionals', *The Philosophical Review* 88, 4 (1979), pp. 544-64, p. 551.)

However, I have an answer to this sort of objection, which is as follows. Sometimes we assert a conditional because its antecedent *entails* its consequent, and this is obviously the case with the conditional 'If Smith strikes that match and at the same time holds it under water, Smith will strike that match'. (I should remark, though, that I don't concede that 'A & B' *always* entails 'A'; in particular, I refuse to concede this where 'A' and 'B' are logically incompatible, e.g. where 'B' = '~A'. I shall return to this matter at the end of the paper.) Other conditionals obviously aren't like this – in particular, 'If Smith strikes that match, it will light' is not. But this, I suggest, shows that we have here two quite different *kinds* of conditional, with quite different kinds of truth-conditions. That being so, however, the only sort of fallacy involved in the application of (T) just mooted is effectively one of *equivocation*. Provided we restrict ourselves, in our applications of (T), to conditionals *all of the same kind*, we shall never be troubled by apparent counterexamples like the one just mentioned. What I am suggesting, then, is that (T) is valid provided the conditional connective '$\rightarrow$' is interpreted *throughout* as expressive *either* of an entailment relation between the various antecedents and consequents *or* of the sort of relation we discern in a conditional like 'If Smith strikes that match, it will light'. My point would then be that while (T*) must of course be accepted as a substitution-instance of (T), it is not legitimate to admit as a particular application of (T*) any case in which a mixture of different kinds of conditional appears. Thus where '$\rightarrow$' in *both* premises is interpreted as expressive of an entailment relation, we may indeed infer '$(A \& B) \rightarrow C$' from '$(A \& B) \rightarrow A$' and '$A \rightarrow C$'; but where '$\rightarrow$' in the first premise is interpreted as expressive of an entailment relation while in the second premise it is *not*, we may *not* infer '$(A \& B) \rightarrow C$' from '$(A \& B) \rightarrow A$' and '$A \rightarrow C$'.

I should just remark that I am not denying that there may well *be* valid arguments involving mixtures of different kinds of conditional, only that their validity can derive from their instantiating a schema like (T). If one wants to write inference-schemata displaying the logical form of such arguments, one must make a *formal* distinction between the different types of conditional figuring therein. Thus, for example, it might seem that the following *is* a valid inference-schema:

$$(T') \quad A \rightarrow B,\ B \Rightarrow C \vdash A \rightarrow C$$

where '$\Rightarrow$' is interpreted as expressive of an entailment relation while '$\rightarrow$' is interpreted as expressive of the sort of relation we discern in a conditional like 'If Smith strikes that match, it will light'. (In fact, however, the validity of (T′) is questioned by, for instance, Donald Nute: see his *Topics in Conditional Logic*, Dordrecht: Reidel, 1980, p. 7.) My contention, of course, would be that whether or not (T′) is valid, the following is certainly *not*:

(T″) $A \Rightarrow B, B \rightarrow C \vdash A \rightarrow C$

I can now return to the objection, mooted earlier, that I have no right to restrict the sense of the modal operator in (W) to 'epistemic' possibility. My point would be that how one interprets the modal operator in (W) should depend on how one interprets the conditional connective '$\rightarrow$' figuring therein. In particular, only when the connective is taken as expressive of an entailment relation should the modal operator be interpreted as expressive of mere logical possibility. (As for (P), we obviously *are* committed to interpreting the modal operator there as expressive of logical possibility because the conditional connective can *only* be taken as expressive of an entailment relation. It follows, of course, that (T*) in conjunction with (P) in fact only leads to (W) under these same interpretations of the operators and connectives; nonetheless, I am prepared, as I have indicated, to accept (W) under the alternative interpretations as well.)

While on the subject of (W), I should perhaps mention another objection to its validity which has been raised with me. If (W) is valid, then so, we must surely concede, is

(W^+) $\Diamond(A \,\&\, B), A \rightarrow C, B \rightarrow C \vdash (A \,\&\, B) \rightarrow C$

But an apparent counterexample to (W^+) is this. Suppose that in my house there is a switch at the bottom of the stairs and a switch at the top of the stairs and either switch can be used to turn a certain light on or off; and suppose that the light is in fact off. Then it might seem correct to assert both 'If the switch at the bottom of the stairs is turned at time t, the light will go on' and 'If the switch at the top of the stairs is turned at time t, the light will go on'; moreover, it might seem that one could assert both of these conditionals while recognizing the (epistemic) possibility that *both* switches might be turned at time t. Yet one would presumably *not* want to assert 'If both the switch at the bottom of the stairs and the switch at the top of the stairs are turned at time t, the light will go on'. My answer is that precisely because of the known connection between the two switches and the known possibility of their being turned simultaneously, it would *not* in fact be correct to assert either of the first two conditionals. For instance, rather than asserting 'If the switch at the bottom of the stairs is turned at time t, the light will go on', one should, in the circumstances hypothesized, assert 'If the switch at the bottom of the stairs is turned at time t *and the switch at the top is not turned at t*, the light will go on'. After all, suppose someone asked me 'Will the light go on if I turn the switch at the bottom of the stairs?'. To answer with an unqualified 'Yes' would be quite misleading: what I should answer is 'Yes – provided no one is turning the switch at the top of the stairs at the same time'. Of course, I *could* quite properly answer with an unqualified 'Yes' if I were reasonably sure that no one else *would* turn the other switch at the same time: but this is contrary to what was hypothesized. (Again we must be careful not to confuse epistemic with mere logical – or indeed physical – possibility.)

I remarked earlier that I refuse to concede that 'A & B' entails 'A' where 'A' and 'B' are logically incompatible. One advantage of such a position is that it provides us with a very straightforward way of undercutting the notorious 'proof' that a contradiction entails anything whatever, because that 'proof' assumes that 'A & ~A' entails 'A'. But surely, it will be protested, 'A & ~A' *must* entail 'A', because from 'A & ~A' we may deductively infer 'A' by the uncontroversial rule of '&-elimination'. My reply is that in my view we may not legitimately infer *anything* from a contradiction, because we may not legitimately take a contradiction as a premise for an argument. (This is not to say that we may not legitimately argue from premises which contradict *each other*, or which are *mutually* inconsistent: to say this would be to deprive ourselves of *reductio ad absurdum* arguments.) Here it may be objected that sometimes it may not be immediately obvious that a premise is self-contradictory, and the fact may only emerge in the course of attempting to argue from it. But when the fact does emerge the proper response is to say that it was a *mistake* to attempt to argue from that premise. This is not of course to deny that we may *learn* by such mistakes. In describing such an attempt as a 'mistake' I am not necessarily implying that it were better if the attempt had never been made, but only that what was attempted (viz. to argue legitimately from the premise in question) turned out to be impossible.

University of Durham,
50 Old Elvet, Durham DH1 3HN

© E. J. LOWE 1985

SIMULTANEOUS CAUSATION AND CAUSAL CHAINS

By DOUGLAS EHRING

A STANDARD objection to the thesis that all causation is simultaneous causation is that this claim is inconsistent with the existence of chains of causally connected, non-simultaneous events (see for example David Hume, *A Treatise of Human Nature*, ed. L. A. Selby-Bigge, Oxford University Press 1888, p. 76, and Richard Taylor, *Action and Purpose*, Prentice-Hall 1966, p. 38). Defenders of universal simultaneous causation have suggested two possible resolutions of this problem. The first consists of directly denying the supposed incompatibility. Michael Dummett develops a version of this strategy ('Can an Effect Precede its Cause?', *Proceedings of the Aristotelian Society*, supplementary volume 28 1954, pp. 28-9; reprinted in his *Truth and Other Enigmas*, Duckworth 1978, pp. 320-1). The second involves denying the existence of chains of causally connected, non-simultaneous events. On this view it is denied that any two events c and e are such that c and e are non-simultaneous, linked by a chain of events, and either c causes e or e causes c. Myles Brand has recently proposed a version of this second line of reply ('Simultaneous Causation', in *Time and Cause*, ed. Peter van Inwagen, Reidel 1980, pp. 137-53).

In this paper I will argue that neither type of defence against the incompatibility problem is plausible.

I. FIRST STRATEGY

In 'Can An Effect Precede Its Cause?' Dummett states the incompatibility problem as follows: 'if causes are contemporaneous with their effects, we are faced with the dilemma which Hume posed: for the cause of the cause will in its turn be simultaneous with the effect, and we shall be unable to trace the causal ancestry of an event back a single instant in time.' (Dummett p. 320) In response to this objection, Dummett claims that although causes operate to bring about their immediate effects without any lapse of time, we are able to trace the causal ancestry of an event back in time without an arbitrary lacuna in our chain of causal explanation. Causal ancestry over time is possible since a cause may initiate a process the continuation of which beyond that point in time requires no further causal explanation. In order to illustrate what he has in mind Dummett cites an example from Newtonian mechanics:

> In Newtonian mechanics the effect of a force is simultaneous with its being exerted; but what requires to be explained by the action of a force is not, as in Aristotelian mechanics, any deviation from a state of rest, but a deviation from uniform motion in a straight line. (Dummett p. 320)

Hence the immediate effect is the 'start' of the process whose 'continuing on' requires no causal explanation. In turn, the terminus of this latter process will cause a new, remote effect. According to Dummett, universal simultaneous causation and temporally extended causal chains are thus compatible: 'Causes are simultaneous with their immediate effects, but precede their remote effects' (Dummett p. 321).

Dummett's attempted resolution of this difficulty rests on the following picture of causal chains:

$$\begin{array}{llllll} d & \text{causes} & b_1 \ldots & b_n & \text{causes} & e \\ t_1 & & t_1 & t_n & & t_n \end{array}$$

The event d immediately causes b_1 where d and b_1 occur simultaneously. In turn b_1 is the first stage of a non-causal process which terminates in b_n and which takes a certain amount of time in order to run to completion. Finally, b_n, the terminus of the process, immediately causes e and there is no temporal gap between b_n and e.

Dummett's proposal, however, faces a fatal difficulty. Either the intermediate stages of the process $b_1 \ldots b_n$ are caused or they are uncaused. If they are caused then each stage of the process is simultaneous with its immediate cause, and we get temporal collapse. If we say that the various stages other than b_1 are uncaused, as Dummett seems to hold, then we must say that b_n, the terminus of the process, is uncaused. We then would have the following set of statements:

(1) d immediately causes b_1
(2) b_n immediately causes e
(3) b_n is uncaused
(4) d is the remote cause of e

Presumably (4) is true in virtue of the fact that d causes b_1. But if b_1 does not cause b_n we cannot say that d causes e since the cause of e, i.e. b_n, is not caused either immediately by d or remotely by d via b_1.

II. SECOND STRATEGY

The second strategy for resolving the problem of incompatibility is simply to deny the existence (and usually, logical possibility) of causal chains of successive events, at least as these chains are ordinarily understood, i.e. as involving non-simultaneous causation. A proponent of this tactic must provide some alternative account of the structure of chains of temporally extended events which at least are ordinarily taken to be causally connected. Brand in 'Simultaneous Causation' develops an analogue notion of a 'non-simultaneous causal chain' which is meant to accomplish the latter without being incompatible with universal simultaneous causation.

Brand distinguishes between a 'causal interaction' and a 'causal process'. A causal interaction consists of one event causing another event. All causal interactions are simultaneous. A causal process is 'a sequence of events such that the initial and terminal events in the sequence, and only these events are involved in causal interactions' (Brand p. 146). Causal processes can be either temporally extended or simultaneous. 'A *causal chain* consists of at least two causal processes such that the terminal event of each process causally interacts with the initial event of another process' (ib.). Some causal chains are temporally extended. It is important to realize, however, that Brand's temporally extended causal chains do not involve non-simultaneous, causally connected events (where two events c and e are causally connected if and only if either c causes e or e causes c). Pictorially represented Brand's temporally extended 'causal chains' look much like Dummett's (but Brand drops the claim that the first member of the series causes the last member):

$$d_1 \ldots \underset{t_1}{d_n} \text{ causes } \underset{t_1}{b_1} \ldots \underset{t_n}{b_n} \text{ causes } \underset{t_n}{e_1} \ldots e_n$$

d_n and b_1 'causally interact' simultaneously as do b_n and e_1. The events in the sequences $d_1, d_2, \ldots, d_n$, $b_1, b_2, \ldots, b_n$, and $e_1, e_2, \ldots, e_n$ are not causally connected. In turn each such sequence is taken to be a complex event, respectively d, b, and e. Finally, Brand unlike Dummet would say that d_1 does *not* cause e_n.

Brand acknowledges that the latter denial does not correspond with our ordinary way of thinking about such event sequences. As ordinarily understood temporally extended causal chains typically involve non-simultaneous, causally connected events. Brand recommends that these situations need reanalysis: '"Jack's going down on the seasaw causes Jill's going up" suggests that there are two causally related events, though as we have seen, this involves many events and three causal processes' (Brand p. 150). The problem of incompatibility is hence resolved by denying the existence of causal chains of successive events as ordinarily understood. A revised notion of a 'causal chain', which is not incompatible with universal simultaneous causation, is then substituted for this abandoned (in principle) way of talking.

Even if we are willing to accept the counter-intuitive rejection of the ordinary picture of causal chains, Brand's theory faces a serious difficulty. Brand's reanalysis of 'causal chains' does not provide a basis for distinguishing causal from non-causal chains involving successive events. This objection has two parts. First, without some account of the part-whole relation for events this distinction cannot be drawn. Second, even with some account of the part-whole relation, this distinction between causal and non-causal chains cannot be sustained on Brand's view.

Consider the following series of events: e_1 —— e_2 —— f_1 —— f_2. Suppose that e_2 causes f_1. e_2 and f_1 are simultaneous. e_1 occurs before e_2, and f_1 occurs before f_2. Does this series qualify as a 'causal chain' in Brand's sense? That would seem to depend upon whether e_1 and e_2 are temporal parts of the same complex event, and similarly for f_1 and f_2. Certainly if e_1 and e_2, and f_1 and f_2 are not temporal parts of two complex events then this series is not a 'causal chain' in Brand's sense. Without some such restriction any number of non-causal chains in which at least two of the events causally interact will qualify as causal chains. Some criterion of membership as a temporal part of a complex event is required to avoid this result. Brand, however, fails to provide such a criterion and goes so far as to say that '... it is not necessary to provide a criterion for dividing events into spatiotemporal parts' (Brand p. 145).

At this point, a proponent of Brand's view might suggest that a criterion will be forthcoming and hence this objection need not concern us. Even if we grant this reply, however, the objection at hand can be deepened as follows: an intuitively plausible criterion for temporally dividing up events will not guarantee that Brand's theory will adequately distinguish between causal and non-causal chains.

Consider the following case: c is a complex event made up of c_1 and c_2. c_1 occurs before c_2, and c_1 and c_2 are not causally related. c_1 and c_2 share in common no causal antecedents. d causes c_1, and c_2 causes e. On Brand's analysis, it would seem to follow that there is a 'causal chain' from d to e since d causes c_1, and c_2 causes e, and c_1 and c_2 are temporal parts of the same complex event. This result, however, is at odds with what we would normally say about such cases. A specific example will show that this implication does not hold. Suppose that c_1 is the event of John's losing his paper money and c_2 is the subsequent event of John's losing his coins. c_1 is caused by d, the event of John's being robbed. c_2 is caused by some event independent of c_1 and d. John's losing his coins (c_2) causes Mary to be amused (e). On Brand's analysis, we must say that there is a causal chain from John's being robbed to Mary's being amused. Clearly we would not, in this case, ordinarily posit a causal chain linking these two events. It is not the case that we allow in general inferences of the following sort: if d causes c_1 and c_1 is a temporal part of f, then if c_2 is a later temporal part of f and c_2 causes e, then there is a causal chain from d to e.

III

Two lines of reply to the incompatibility objection have been considered. The first attempts to show that the incompatibility is merely apparent. As we have seen, this approach as developed by Dummett faces a dilemma of either temporal collapse or causal lacuna. The second admits the incompatibility of universal simultaneous causation with temporally extended causal chains as ordinarily understood, but denies the existence of such chains, and provides an analogue for the relevant notion. The second approach, however, cannot adequately distinguish between causal and non-causal chains.

Southern Methodist University,
Dallas, Texas 75275, U.S.A.

© Douglas Ehring 1985

LOSING TRACK OF THE SCEPTIC

By ANTHONY L. BRUECKNER

IN 'On the Tracks of the Sceptic' (ANALYSIS 44.3, June 1984, pp. 121-6), Jonathan Dancy criticizes Nozick's anti-sceptical strategy, according to which I know that I am now in a library room even though I do not know that I am not a brain in a vat. This epistemic situation is made possible, says Nozick, by the fact that knowledge is *not* closed under known logical implication. Dancy's main criticism of Nozick's attack on closure rests upon a number of mistaken claims about an epistemic Universalizability Principle and its relation to closure.

Dancy rightly points out that on Nozick's (fairly standard) presentation of the sceptical argument, the sceptic requires both closure and the premise that I do not know that I am not a brain in a vat. Dancy also rightly maintains that the sceptic must offer some reason for accepting the latter premise. Dancy then claims that the premise is justifiable by a further sceptical argument – the 'Argument from Error' – which is founded on a Universalizability Principle (UP) for knowledge analogous to such a principle in ethics. UP is at one point stated as follows: '... if you want to claim that you know now, you must be able to point to some recognizable difference between this case and other similar cases in which you (now) agree that you did not know' (p. 124). Truth, Dancy continues, does not count as a 'recognizable property'. Here is how Dancy uses the principle in order to justify one of the sceptic's premises:

> You do not know you are not a brain in a vat because there is no difference which you can point to between your present situation and the imaginably similar situation in which you are a brain in a vat. In the latter situation you don't know that you are not a brain in a vat, trivially, and hence in the former you don't either (p. 124).

The first thing to note about UP is that it is false if there is a justification condition for knowledge which is independent of the truth condition. If so, then there will be a case where S justifiably believes the false proposition that P. Consider a case in which S has an exactly similar justification for his *true* belief that P. Setting aside Gettier worries (we could assume that a fourth condition for knowing is satisfied in each case), S knows that P in the second case even though he does not know that P in the first case. But UP has the consequence that S does *not* know in the second case since (1) he does not know in the first case, and (2) there is no 'recognizable difference' between the two cases (S has exactly similar justification in each case).[1]

---

[1] For further discussion of principles like UP, see my 'Epistemic Universalizability Principles', *Philosophical Studies* 46, November 1984, pp. 297-305.

The second thing to note about UP is that if closure fails in the cases in which Nozick claims failure, then UP is thereby refuted. This is important because Dancy maintains that even if Nozick is right and closure fails, apparently sinking the sceptical argument from closure, 'there remains a different sceptical argument which establishes that we do not know that we are not brains in a vat', viz. the Argument from Error founded on UP (p. 126). According to Dancy, this argument is 'far more powerful' than the argument from closure and is required in order to justify a premise of the latter argument (that I do not know that I am not a brain in a vat).[2] Thus Dancy continues: 'But if there is such an argument [the 'far more powerful' Argument from Error], disproving ... [closure] cannot be a complete reply to the sceptic ... The spectre of scepticism cannot be removed by disproving one principle of inference' (p. 126). To see that Dancy is mistaken here, consider first that Nozick contends that I know that, e.g., I am now in a library room even though I do not know that I am not a brain in a vat, contrary to closure. Suppose Nozick is correct. Then UP is thereby refuted, since UP not only has the consequence that I do not know that I am not a brain in a vat (as Dancy points out), but also the consequence that I do not know that I am now in a library room. Just as I can point to no recognizable difference between my present situation and one in which I fail to know that I am not a brain in a vat in virtue of the fact that I *am* a brain in a vat, similarly I can point to no recognizable difference between my present situation and one in which I fail to know that I am in a library room in virtue of the fact that I am in a vat and not in a library room. UP, just like closure, has the consequence that my claim to know that P stands or falls with my claim to know that I am not a brain in a vat deceived into mistakenly believing that P.

I conclude that even if one (implausibly) maintains that UP is a compelling principle which can generate scepticism, one should nevertheless recognize that it is *not* a principle which is independent of closure in the way Dancy claims. If closure fails, then so does UP.

Yale University,
New Haven, Connecticut 06520,
U.S.A.

© ANTHONY L. BRUECKNER 1985

---

[2] Even though Nozick's externalist analysis of knowledge in terms of subjunctives explains why I do not know that I am not a brain in a vat (one of the subjunctive conditions is violated), Dancy claims that Nozick still needs to appeal to an argument from UP in order to provide 'independent support' for his view that I do not know that I am not a brain in a vat. But it is completely unclear why an externalist should be required to appeal to an *internalist* argument in order to justify a component of his epistemological position, which component is justified anyway on externalist grounds (violation of a subjunctive condition). For Dancy's discussion, see 125-6.

DEREK PARFIT AND GRETA GARBO

By GEOFFREY MADELL

IN chapter 11 of his *Reasons and Persons* (Clarendon Press, 1984) Derek Parfit develops what he regards as a central argument against the Non-Reductionist view of personal identity. This is the view that personal identity is strict, and not analysable in terms of physical and/or psychological continuity. On this view personal identity involves, in Parfit's words, a 'further fact' in addition to any such continuity, even a 'separately existing entity', such as a Cartesian ego. Parfit's Reductionist view, by contrast, is just that personal identity through time involves nothing more than such continuities, and of these psychological continuity and connectedness is much the more important. On this view, what we call (however loosely) 'personal identity' is essentially a matter of *degree.*

The crux of Parfit's case against the Non-Reductionist view can be briefly stated: hard though it may be to believe in the Reductionist position, there is just no *evidence* that personal identity involves anything other than physical and/or psychological continuity, and plenty of evidence against such a supposition. I shall argue that this approach is wholly misguided. Not only has this approach resulted in Parfit's failure to get to grips with the powerful *logical* case for a non-reductionist position (though Parfit does express the hope of being able to discuss the arguments of writers such as Swinburne and myself elsewhere), but it also reveals a fundamental confusion, both about what evidence (if any) we might actually have in this area, and also about how any such (putative) evidence is to be *interpreted.*

What is the evidence to which Parfit draws our attention? Curiously, instead of anything we would recognize as hard scientific evidence, Parfit offers us a speculation, concerning what he calls the Combined Spectrum. This spectrum embraces 'all of the possible variations in the degrees of *both* physical *and* psychological connectedness' (p. 236). At the near end of the spectrum is the normal case in which a future person would be fully continuous with, say, Derek Parfit, as he is now, both physically and psychologically. At the far end of the spectrum is the case where scientists would destroy Parfit's brain and body, and then create, out of new organic matter, a perfect replica of someone else, say, Greta Garbo. In between these two poles is a whole range of cases, in which a progressively greater physical-cum-psychological part of Derek Parfit is replaced by a progressively greater part of the Garbo replica. It is clear that there is a whole range of cases in the middle of this spectrum where there is no clear answer to the question, is this person Derek Parfit or Greta Garbo? A further point is that the spectrum, in combining *both* psychological *and* physical aspects, blocks any hope of grounding the supposed all-or-nothing character of personal identity in, say, physical continuity; this is an escape-route which would have appeared open had the spectrum embraced psychological continuity only, and is a route which, as Parfit points out, Bernard Williams has attempted to take.

Anyone who believes that personal identity is all-or-nothing must hold that somewhere in this spectrum there is a sharp borderline, on the one side of which we have Derek Parfit, and on the other Greta Garbo. But it is difficult to believe in the existence of such a borderline. How could the replacement of a few extra cells have such a dramatic outcome? However, were there to be some critical percentage of cells whose replacement, we discover, utterly destroys psychological continuity, we should then have an empirical demonstration of the strictness of personal identity. Parfit even allows that this would be evidence for the existence of something like a Cartesian ego (p. 237). But there is in fact no such evidence. There is, on the other hand, much evidence that the carrier of psychological continuity is the brain, and, consequently, every reason to believe that psychological connectedness could hold to any degree. The possibility of such a Combined Spectrum thus provides, Parfit claims, a strong argument for the Reductionist view of personal identity.

The Combined Spectrum argument fails. It fails, first, because it is impossible to accept this speculation as presenting any sort of evidence at all. If, as Parfit believes, the issue between opposing views of the nature of personal identity is to be settled by considering the evidence for each view, what possible contribution can a mere *speculation* make to this assessment? It is important to realize that the purpose of this speculation is not to reveal anything about the *logic* of personal identity, but to show us what persons *in fact* are. Persons, or so we are meant to conclude after considering the Combined Spectrum, are composite, capable of being dismantled and re-assembled rather like tables and chairs. If any of this is impossible, the impossibility is 'merely technical'; nothing in the operations envisaged is contrary to the laws of nature.

How, one must ask, can Parfit know this? Parfit actually admits at one point that 'it seems likely that it would never be possible to divide the lower brain in a way which did not impair its functioning (p. 225), and this does seem to be an admission that the operations suggested in speculation might, after all, be contrary to the laws of nature. This moment apart, however, Parfit sees no problem. The Combined Spectrum case is just a rather dramatic way of setting out what we already know, or, at least, the implications of what we already know. That this is what Parfit thinks becomes clear on page 228, where Parfit tells us that 'since our psychological features depend on the states of our brains, these imagined cases are only technically impossible'. Persons, then, are such that it is in principle possible to take out a chunk from one person's mind and replace it by a chunk from someone else's mind, and we know this because we know that the mental is dependent on the physical.

How can Parfit have supposed that his claim can be supported in this way? No one would deny that 'our psychological features depend on the states of our brains'; even the interactionist dualist and believer in libertarian free-will might well accept that in this world it is a necessary condition of any particular psychological state's being realized that some particular brain state be also realized. But even if we take the claim about the dependence of the mental on the physical as the physical determinist takes it – that every mental state and event is fully determined by states of and events in our brains – it is still clear that this offers no support for Parfit's view of psychological states as in principle extractable and replaceable by others from other persons. Even the physical determinist might well regard the suggestion that, say, Parfit's interest in Buddhism could be taken out of Parfit's mind and replaced by Greta Garbo's Longing to be Alone as nothing but a fantasy.

Suppose, however, we were to accept the claim that, just because our psychological features depend on the states of our brains, it follows that chunks from different minds could be somehow welded together. Even granted this, it is quite unclear that the outcome must be as Parfit supposes. Parfit takes it that if we could in this sort of way bring together some of the psychological features of Derek Parfit with some of Greta Garbo's, we should then have a person who is neither Derek Parfit nor Greta Garbo; and as we move towards one or other end of the Combined Spectrum, so we find persons who are more Parfit than Garbo, and *vice versa*. This way of looking at things is meant to support an essentially *impersonal* view of experiences.

But, not only is there nothing about Parfit's speculation which compels us to construe it in this way, it is not even, to my mind, the most natural and obvious way of construing it. It seems more natural to suppose that were it possible to combine some of your mind with some of mine, what we would have is something like two minds in the one skin, a pretty startling case of multiple personality, perhaps. At any rate, we need solid argument to support the claim that, when some of my psychological features are combined with psychological features from someone else, these features cease to be *mine*. And if the outcome of the operations were to be as I am suggesting, it is fairly clear that this would not support Parfit's reductionist view at all. If Parfit and Garbo remain distinct personalities in the one skin, then there is no person who, being a judicious mix of both, is identical with neither, and Parfit's claim that personal identity through time is a matter of degree can receive no backing from this quarter. The fundamental point is that we haven't the slightest reason to suppose that the Combined Spectrum case is empirically possible, or that, if it were possible, the outcome would be as Parfit supposes. In brief, the claim that only the Reductionist view can accommodate the available data collapses if, as is the case, there is no such body of data.

The second reason why the Combined Spectrum argument fails is that, even if everything were to happen according to Parfit's plan, and what results from the operations in the middle of the Spectrum is a series of persons, each with a tolerably unified consciousness comprising some of Parfit's and some of Garbo's mental features, it still would not follow that we had any sort of argument for the Reductionist position. Parfit too readily supposes that, were he to lose some of his own psychological features and acquire some of Garbo's, the resulting person would be some sort of mixture of Parfit and Garbo. But, far from this being the only possible way of construing this sort of change, such an interpretation is in fact highly problematic, and actually presents a particular problem for Parfit. The reason why the interpretation is problematic is this: while I can easily imagine losing some particular psychological characteristic and at the same time acquiring, say, an interest in Buddhism and a longing to be alone, it seems obvious that what I am imagining is *myself*, albeit changed in these rather interesting ways, rather than a new person comprising bits of Madell, Parfit and Garbo. Even more clearly, if I suppose that I might now have been different in this sort of way, it is *my* consciousness I imagine to be now different in these respects, not that of some composite person. From the first person viewpoint, the latter isn't a conceivable possibility at all; I cannot make sense of the suggestion that there is a possible world in which there is a present state of consciousness which is only partly mine. There are complexities here which I cannot discuss in a short paper, but it is clear that there is a crucial first-person dimension to this issue, and that Parfit's failure to discuss it is an instance of the general disregard of the first-person perspective which, in *The Identity of the Self* (Edinburgh University Press, 1981), I argued was at the root of so much that is unsatisfactory about contemporary discussions of personal identity.

One might reply that the crucial factor is not just the change in psychological make-up but the fact that this change is brought about by some sort of physical transplant – the taking of cells from replicas of Parfit and Garbo respectively and their insertion into my brain. Such a reply, in insisting on the importance of the particular physical cause of the psychological change, would run clean counter to Parfit's own repeated claim that the particular cause of psychological continuity is quite irrelevant; all that matters is that the cause is reliable. It is difficult to see why this claim about causality should be dropped when the issue is psychological *change* as opposed to *continuity*. But even if such a move were thought acceptable, the problem of reconciling this reductionist conclusion with our first-person intuitions would remain. Nothing about the Combined Spectrum case shows us how to deal with such intuitions, and certainly nothing about it gives us any ground for disregarding them.

The Reductionist view, Parfit acknowledges, is 'hard to believe'. In particular, it is desperately hard to reconcile the reductionist belief that the essence of personal identity through time is psychological continuity and connectedness with our ordinary attitude to pain. How, Parfit asks (p. 230), could a loss of memory and a total change of character occurring while one is in agony mean that one ceased to exist half-way through the agony? Yet, strong though this consideration is, the Combined Spectrum argument compels us to reject it, so Parfit argues, for it shows that all the evidence supports the Reductionist position. I think it is clear that Parfit regards the Combined Spectrum argument as overriding not only this but *every* consideration that might be brought in support of a non-reductionist position. The truth, however, is that the Combined Spectrum case is just a speculation, and there is not the slightest ground for supposing that it describes an empirical possibility. Moreover, even if there were the sort of data that Parfit envisages, far from this justifying a rejection of all non-reductionist considerations, it is in fact just those considerations which make it extraordinarily difficult to interpret the data as Parfit would like.

University of Edinburgh,
David Hume Tower,
George Square, Edinburgh EH8 9JX

© GEOFFREY MADELL 1985

THE UMPIRE'S DILEMMA

By COLIN RADFORD

THE laws of cricket (30-9) state that a batsman may be out in any one of the following ways: bowled, timed out, caught, handled the ball, hit the ball twice, hit wicket, lbw (leg before wicket), obstructing the field, run out, and stumped. There is also a time-honoured and very powerful convention, and a rather specific law (27.4) which seems to imply, that in giving a batsman out the umpire must be in no doubt that the batsman is out; any doubt in his mind must benefit the batsman rather than the fielding side. (The laws require in effect that the umpire be certain that the batsman is out (and how). They do not further specify that this certainty be justified, or that the umpire should believe it to be justified, as some 'analyses' of knowledge specify. But this is, presumably, because such a requirement would be otiose — it goes without saying that the good, i.e. informed, able, conscientious, and unbiased umpire's certainty is justified.)

Now envisage the following situation: the bowler delivers a ball which strikes the batsman on his pads and then carries through to the wicket-keeper, who catches it. Bowler and wicket-keeper appeal 'How's that?' The difficulty for the umpire is that he is not sure whether the ball made contact with the batsman's bat before striking his pads. If it did, the batsman is out caught; if it did not, he is in this case out lbw. So it seems clear that he is out one way or other, the only difficulty is which. But since the umpire is uncertain that the ball struck the batsman's pads before striking his bat (if it struck it at all), he cannot give him out lbw. And since by the same token he is uncertain that it did strike the batsman's bat, he cannot give him out caught either. What then should the umpire do?

If he is a robust realist, he may well say, 'The batsman *is* out one way or the other, lbw or caught, and it doesn't really matter which'. But there is no way of being out 'lbw or caught', a batsman has to be out one way or another. To get over this difficulty the realistic umpire may ask the batsman if the ball struck his bat and, if it did, whether it did so before striking his pads. But the batsman may not know or, as he is entitled to do, he may refuse to answer.

This difficulty may lead one to sympathize with the umpire who is legalistic in attitude, or even a philosophical intuitionist. He may say that since he cannot give the batsman out either lbw or caught, and since he cannot give him out disjunctively, i.e., lbw—or—caught, he cannot and should not give him out. Yet this seems most unfair to the fielding side because, surely, the batsman *was* out lbw or caught, even though the umpire cannot say which. (To give him not out is to give him the benefit of the doubt twice. This would appear to be an error of generosity parallel but opposite to that of putting someone in double jeopardy.) What then should the umpire decide? Should the rules of cricket be amended?

A further question is this: we have a very clear distinction in our minds when we say that the batsman is out lbw or caught, but he cannot be out 'lbw—or—caught', i.e., he must be out one way – or the other! What is clear in our minds seems to issue in something that avoids contradiction only by the dubious – because *ad hoc* – use of hyphens and 'scare' quotes. We can avoid this difficulty by saying that what is to be written in the score book is 'lbw' or 'caught', not 'lbw or caught'. But this point now made clear, the second umpire's decision rests on another distinction, viz. between the umpire's *knowing* that the batsman is out lbw or caught and his *knowing* that he is out lbw or that he is out caught. That is to say in symbolic notation, the legalistic or intuitionist umpire claims that from Uk(p v q) it does not follow that Ukp v Ukq.

Yet this claim contradicts the claim usually attributed to the intuitionist, viz., that if one knows that p or q then one *does* know that p or one knows that q. That is, unless one can construct a proof of p or of q one does not know that p or q and it is *not true* that p or q.

What this means is that if the second umpire *is* an intuitionist rather than a mere legalist, he will not merely refuse to give the batsman out but will deny that the batsman was in fact out one way or the other. But surely this is counter-intuitive and more legalistic than the legalist's position, which would be that he cannot *say* which way the batsman was out and so cannot *give* him out. Intuitionism would deny the existence of a problem.

Does this then leave the legalistic umpire's decision as the correct one? And what light, if any, does this case throw on the nature and merits of intuitionism, or cricket?

University of Kent,
Keynes College,
Canterbury, Kent CT2 7NP

© COLIN RADFORD 1985

REASONS, DILEMMAS AND THE LOGIC OF 'OUGHT'

By CASEY SWANK

CAN dilemmas arise? Most anyone you ask will tell you that he has faced quite a few of them. Still, arguments designed to show that dilemmas cannot arise *have* been advanced – and two of these are *prima facie* sound. (Terrance C. McConnell discusses variants of both of these, in his 'Moral Dilemmas And Consistency In Ethics', *Canadian Journal of Philosophy* vol. 8 no. 2 (June 1978), pp. 269-87. The first argument is better known,and is further discussed in E. J. Lemmon's 'Moral Dilemmas', *The Philosophical Review* vol. 71 (April 1962), pp. 139-58, and B. A. O. Williams' 'Ethical Consistency (I)', *Proceedings Of The Aristotelian Society* supp. vol. 39 (10 July 1965), pp. 103-24.) I mean to show that neither of these arguments is in fact sound.

The arguments in question both begin with the observation that:

(1) I do not face a dilemma vis-à-vis A and B unless:

(i) I can do A,
(ii) I can do B,
(iii) I cannot do A-and-B,
(iv) I ought to do A, and
(v) I ought to do B.

The first argument then proceeds as follows:

(2) Given (iv) and (v), it follows that:

(vi) I ought to do A-and-B.

(3) Given (vi), though, it follows that:

(vii) I can do A-and-B.

Evidently, then, I do not face a dilemma vis-à-vis A and B unless, *per impossibile*, I both can and cannot do A-and-B, i.e. unless conditions (iii) and (vii) jointly obtain. We must therefore conclude that dilemmas cannot arise after all.

Letting 'DXY' stand for 'I face a dilemma vis-à-vis X and Y', 'OX' stand for 'I ought to do X', and 'CX' stand for 'I can do X', this first argument may be stated as follows:

(1) $DAB \supset (((CA \mathbin{\&} CB) \mathbin{\&} {\sim}C(A \mathbin{\&} B)) \mathbin{\&} (OA \mathbin{\&} OB))$
(2) $(OA \mathbin{\&} OB) \supset O(A \mathbin{\&} B)$
(3) $O(A \mathbin{\&} B) \supset C(A \mathbin{\&} B)$
∴ (4) ${\sim}DAB$

This argument *looks* sound. But is it really? I mean to show that it is not. First things first, though: let us have a look at that second argument.

Beginning, as the first argument did, with the observation that:

(1) I do not face a dilemma vis-à-vis A and B unless:

(i) I can do A,
(ii) I can do B,
(iii) I cannot do A-and-B,
(iv) I ought to do A, and
(v) I ought to do B,

the second argument then proceeds as follows:

(5) Given (ii), (iii) and (iv), it follows that:

(viii) I ought not to do B.

(6) Given (viii), though, it follows that:

(ix) It is not the case that I ought to do B.

Evidently, then, I do not face a dilemma vis-à-vis A and B unless, *per impossibile*, it both is and is not the case that I ought to do B, i.e. unless conditions (v) and (ix) jointly obtain. And so we must (again) conclude that dilemmas cannot arise after all.

Sticking with the notation adopted above, this second argument may be stated as follows:

(1) $DAB \supset (((CA \mathbin{\&} CB) \mathbin{\&} {\sim}C(A \mathbin{\&} B)) \mathbin{\&} (OA \mathbin{\&} OB))$
(5) $((CB \mathbin{\&} {\sim}C(A \mathbin{\&} B)) \mathbin{\&} OA) \supset O{\sim}B$
(6) $O{\sim}B \supset {\sim}OB$
∴ (4) ${\sim}DAB$

This argument is at least as compelling as the first one. But is either truly sound? We have to ask: what, specifically, is meant here by 'I ought to do X'?

When I say that I ought to do something, I sometimes just mean that the reasons for me to do it are *weighty*. What I usually mean, though, is that the reasons for me to do it *outweigh* those for me *not* to do it. 'I ought to do X' may thus be defined in either of two ways:

(a) 'I ought to do X' = df 'The reasons for me to do X are weighty'
(b) 'I ought to do X' = df 'The reasons for me to do X outweigh those for me not to do X'

Letting 'RX' stand for 'the reasons for me to do X', 'XW' stand for 'X are weighty', and '$X > Y$' stand for 'X outweigh Y', (a) and (b) may be stated as follows:

(a) $\text{'}OX\text{'} = \text{df '}RXW\text{'}$
(b) $\text{'}OX\text{'} = \text{df '}RX > R{\sim}X\text{'}$

(Distinctions *roughly* resembling that between (a) and (b) have been drawn by: Philippa Foot, in 'Moral Realism And Moral Dilemma', *The Journal Of Philosophy* vol. 80 no. 7 (July 1983), pp. 379-98; Roger Trigg, in 'Moral Conflict', *Mind* vol. 80 no. 317 (January 1971), pp. 41-55; and Bas C. Van Fraassen, in 'Values And The Heart's Command', *The Journal Of Philosophy* vol. 70 no. 1 (January 1973), pp. 5-19.) The question is: do the proponents of the arguments in question intend (a) or do they intend (b)? Let us consider both possibilities.

Given (a), the first argument goes like this:

(1a) $DAB \supset (((CA \,\&\, CB) \,\&\, {\sim}C(A \,\&\, B)) \,\&\, (RAW \,\&\, RBW))$
(2a) $(RAW \,\&\, RBW) \supset R(A \,\&\, B)\, W$
(3a) $R(A \,\&\, B)\, W \supset C(A \,\&\, B)$
$\therefore$ (4) ${\sim}DAB$

and the second one goes like this:

(1a) $DAB \supset (((CA \,\&\, CB) \,\&\, {\sim}C(A \,\&\, B)) \,\&\, (RAW \,\&\, RBW))$
(5a) $((CB \,\&\, {\sim}C(A \,\&\, B)) \,\&\, RAW) \supset R{\sim}BW$
(6a) $R{\sim}BW \supset {\sim}RBW$
$\therefore$ (4) ${\sim}DAB$

Is either of these arguments sound?

Imagine: I have been given the day off, and I may spend it with whomever I like – my family, my friends or both. My family and I stand to gain a lot from a day spent together, and so do my friends and I: the reasons for me to spend the day with my family (RA) are weighty, and so are those for me to spend it with my friends (RB). Were (2a) true, then, it would follow that the reasons for me to spend the day with my family *and* my friends (R(A & B)) are also weighty. In fact, though, they are not: each of the parties involved would find such a day ill-spent. Evidently, then, (2a) is false. If follows that, given (a), the first argument is unsound. So what about the second one?

Imagine: The doctor has advised that I take the day off, so as to get some much-needed rest: the reasons for me to take the day off ($R{\sim}B$) are weighty. Were (6a) true, then, it would follow that the reasons for me *not* to take the day off (RB) are *not* weighty. In fact, though, they *are* weighty: were I to take the day off, my week's earnings would be insufficient to accommodate my family's needs. Evidently, then, (6a) is false. It follows that, given (a), the second argument is also unsound.

Given (a), neither of the arguments in question is sound. I doubt, however, whether the proponents of these arguments intend (a). To begin with, (2a) and (6a) are both so *obviously* false that no one seriously intent on proving anything would employ either of them to that end. Moreover, as I have observed already, it is only atypically that I intend (a), in saying that I ought to do something: I *usually* intend (b). Finally, the proponents of these arguments use 'I ought to do X' where they intend 'I ought, *all things considered*, to do X' – and the sense of this latter phrase is not so well captured by (a) as it is by (b). (On this, see Richard Fumerton's 'Reasons And Value Judgments', *The Journal Of Value Inquiry* vol. 13 no. 4 (Winter 1979), pp. 259–73.) So let us see what happens to the arguments when (b) is plugged in.

Given (b), the first argument goes like this:

(1b) $DAB \supset (((CA \,\&\, CB) \,\&\, {\sim}C(A \,\&\, B)) \,\&\, (RA > R{\sim}A \,\&\, RB > R{\sim}B))$
(2b) $(RA > R{\sim}A \,\&\, RB > R{\sim}B) \supset R(A \,\&\, B) > R{\sim}(A \,\&\, B)$
(3b) $R(A \,\&\, B) > R{\sim}(A \,\&\, B) \supset C(A \,\&\, B)$
∴ (4) ${\sim}DAB$

and the second one goes like this:

(1b) $DAB \supset (((CA \,\&\, CB) \,\&\, {\sim}C(A \,\&\, B)) \,\&\, (RA > R{\sim}A \,\&\, RB > R{\sim}B))$
(5b) $((CB \,\&\, {\sim}C(A \,\&\, B)) \,\&\, RA > R{\sim}A) \supset R{\sim}B > RB$
(6b) $R{\sim}B > RB \supset {\sim}RB > R{\sim}B$
∴ (4) ${\sim}DAB$

Neither of these arguments is so patently fallacious now as it was before. But is either yet sound?

We should be anxious, first, to examine (2b) and (6b) – as their (a)-counterparts each proved erroneous. (2b) *looks* true to me: unable to see anything wrong with it, I shall approve it without further ado. As for (6b), he who would question it will find himself having to explain how it is that the reasons for me to do something can outweigh themselves (since $(R{\sim}B > RB \,\&\, RB > R{\sim}B) \supset RB > RB$)! It begins to appear as though these arguments may be sound after all. But what about the remaining premises – (3b), (5b) and (1b)?

According to (3b), the reasons for me to do A-and-B do not outweigh those for me not to do A-and-B unless I can do A-and-B. Is this indeed true? I think so. Here is why. A reason for me to do something *recommends* that I do that thing. But a recommendation that I do something cannot be efficacious – cannot carry any *weight* – unless I can, but need not, do the thing in question. It follows that:

(p) The reasons for me to do X do not have any weight unless I can, but need not, do X.

Letting '$X > 0$' stand for 'X have some weight', (p) may be stated as follows:

(p) $(X)(RX > 0 \supset (CX \,\&\, C{\sim}X))$

Given (p), though, it follows that (3b) is true – for the reasons for me to do A-and-B cannot outweigh those for me not to do A-and-B unless the former have *some* weight, in which case, (p) informs us, I can do A-and-B. (N.B. (p) also gives us (3*a*).) And so the proponent of the first argument advances to third base. Let us next see whether his teammate may join him there, i.e. let us have a look at (5b).

Imagine that I can do B only by not doing A, and that the reasons for me to do A outweigh those for me not to do A. Does it follow, as it would were (5b) true, that the reasons for me not to do B outweigh those for me *to* do B? I think so. Here is why. If the only way that I can do something is by not doing something else, any reason for me to do the latter thing is – in effect – a reason for me *not* to do the former. It follows that:

(q) If I can do Y only by not doing X, then the reasons for me not to do Y include, and so are at least as weighty as, those for me to do X.

Letting '$X \geqslant Y$' stand for 'X are at least as weighty as Y', (q) may be stated as follows:

(q) $(X)(Y)((CY \,\&\, {\sim}C(X \,\&\, Y)) \supset R{\sim}Y \geqslant RX)$

Given both (p) and (q), though, it follows that (5b) is true. (N.B. (q) also gives us (5*a*) – and without the aid of (p).)

Evidently, then, (2b), (3b), (5b) and (6b) are each true. (And so it appears all the more likely that the proponents of the arguments in question intend (b).) So each of the arguments in question is sound, given (b), if and only if (1b) is true. Is it, though? I mean to show that it is not.

According to (1b), I do not face a dilemma vis-à-vis A and B unless, among other things:

(ii) I can do B,
(iii) I cannot do A-and-B, and

(x) The reasons for me to do B outweigh those for me not to do B.

Given (ii) and (iii), (q) informs us that:

(xi) The reasons for me not to do B are at least as weighty as those for me to do A.

Given (x) and (xi), though, it follows that:

(xii) The reasons for me to do B outweigh those for me to do A.

So (1b) tells us that (xii) is among the conditions that must obtain for me to be faced with a dilemma vis-à-vis A and B. But is it really?

I offer for your appraisal the following alternative account of what would be the case were I ever to land myself in a dilemma:

(D) I face a dilemma vis-à-vis A and B if and only if:
  (i) I can do A,
  (ii) I can do B,
  (iii) I cannot do A-and-B,
  (xiii) The reasons for me to do A are weighty,
  (xiv) The reasons for me to do A are neither weightier than nor outweighed by those for me to do B, and
  (xv) If I can do X only by not doing A *and* not doing B, then the reasons for me to do X are outweighed by those for me to do A.

(N.B. (D) gives us (1*a*). And so we may well suspect that the proponents of the arguments in question have unwittingly exploited the ambiguity of 'I ought to do X'.) Given (D) – and I have yet to hear of a counterexample to it – (1b) says that the conditions enumerated in (D) cannot jointly obtain unless (xii) does, too. Imagine, then, that I can, but need not, do A, and that the reasons for me to do A are weighty, but neither weightier than nor outweighed by those for me not to do A. Does it follow, as we now see that it would were (1b) true, that the reasons for me to do A actually *are* outweighed by those for me not to do A? Try as I might, I cannot see how. (Can *you*?) I have therefore to conclude that (1b) is false, and hence that, given (b), neither of the arguments in question is sound.

In reply to all of this, the proponent of one of the arguments in question may simply insist that he intends neither (a) nor (b). So what *does* he intend? What is wanted (for his purposes) is some definition of 'I ought to do X' that at once gives us (1), (2) and (3) – or (1), (5) and (6). Is there such a definition? I seriously doubt it. If there is, though, I am anxious to hear of it.[1]

University of North Carolina,
Chapel Hill, North Carolina 27514, U.S.A.

© CASEY SWANK 1985

---

[1] For their help with this paper, I wish to thank my wife Mary, William Lycan, and (especially) Laurence Thomas – none of whom accepts any of the falsehoods I have advanced.

A TRANSPARENT CASE FOR SUBJECTIVISM?

By C. L. HARDIN

IN *The Subjective View* (Oxford, the Clarendon Press, 1983, pp. 35–7), Colin McGinn maintains that 'the impossibility of transparent white ... is just that nothing could *look* transparent and white'. He goes on to maintain that something's looking white cannot be explicated in objectivist terms, i.e. terms which refer only to the properties and states of physical objects. He argues for his account by rejecting an alternative, objectivist analysis, which he expresses in this way: 'For a surface to be transparent is for it to let through the incident light, whereas for it to be white is for it to reflect all of the light; so transparent white is impossible because a surface cannot simultaneously interact in both these ways with the light incident upon it.'

Here is McGinn's argument against this plausible analysis:

> But now consider the case of a Martian whose visual apparatus is so constructed that he has an experience as of something white when light from what *we* call the green part of the spectrum enters his eye. Given the relativity of ascriptions of colour, it seems that we must allow that such surfaces are white for Martians, i.e. they speak truly when they ascribe whiteness to these (green for us) surfaces. But then, according to the physical explanation just suggested, the Martian ought to be able to envisage the possibility of a transparent white surface; it ought to be just as conceivable to him as a transparent green surface is to us. But of course this just contradicts the acknowledged necessity of the non-existence of transparent white surfaces. I think that this shows that the possibility of transparent white cannot be explained in this objective way.

We shall see that the objectivist analysis with which McGinn has presented us is indeed deficient. But, I shall argue, this shortcoming may be remedied entirely within the objectivist framework, although it will require more than an appeal to the physical facts alone. When this is done, it will also be clear that McGinn's Martian example is without force against objectivism, and that it will certainly not support the moral which he draws from it:

> As Wittgenstein says, though there are certain connections between the laws of appearance and physical laws, the former cannot be explained or analysed in terms of the latter. The proper explanation is that in coming to be a conscious perceiver you come to conform to a rich system of phenomenological laws governing how things can seem, these laws arising from the very nature of perceptual experience.

The puzzle about the non-existence of transparent white may easily be transposed into a puzzle about mirrors: there can be green mirrors, red mirrors and even gray mirrors. Why are there no white mirrors? Since white objects are objects which reflect a large percentage of the incident light nonselectively, why isn't a highly reflective nonselective mirror a white mirror?

Think about objects we do call white, such as a snowball or an ermine coat. Unlike the mirror, they do not reflect the light coherently and evenly, so as to permit the formation of images. Their irregular surfaces diffuse the light so that the rays are scattered randomly rather than being reflected in an orderly way. Smooth and polish the surface of a white object and its whiteness disappears as the surface comes to be capable of reflecting an image. On the other hand, perform the same operations on a suitable green object, and you can make the surface mirrorlike without a reciprocal disappearance of greenness. What is true of green is true as well of the other chromatic colours – red, blue and yellow. It is also true of gray: a mirror can be gray, but never white.

Diffusely reflecting gray things, on the other hand, belong in the same colour family as black and white. Together they comprise the hueless, or achromatic, colours. For diffusely reflecting surfaces, white and black are at opposite ends of a single achromatic series, the members of which differ only in reflectance and, consequently, lightness. In such a series white is the upper lightness limit of gray. By contrast, the upper lightness limit of a series of gray mirrors is not, we usually say, a white but a clear mirror. Similarly, the upper lightness limit of a sequence of transparent grays is a clear transparency, while the upper limit of a sequence of diffusely transmitting (translucent) grays is white.

In the case of the achromatic colours, then, we most often choose to call the upperlightness limit 'white' for light diffusers and 'clear' for non-diffusers. It is important to see that we can also choose *not* to distinguish the two cases. Indeed, we do use 'white' to refer to some clear substances – white wine, for example. One of the O.E.D. entries for the adjective 'white' is: 'colourless or uncoloured, as glass or other transparent substances'. However, 'white' is paradigmatically used to refer to light diffusing substances; the first O.E.D. 'white' entry is: 'of the colour of snow or milk'.

Obviously the relevant difference between a nonselective plane mirror and a snowball is an appearance difference, but it is not one which is subjective in any deep sense. It is, rather, directly related to the optical characteristics of surfaces. A near-perfect plane mirror carries no visible marks of its character, for it alters only the direction of the light incident upon it. It is thus uncoloured, just as a loudspeaker is said to be uncoloured if it does not noticeably add to or subtract from the structure of its input. A very good mirror is the best bearer of visual illusion because of its perfect visual self-effacement, its lack of any colour peculiar to itself. A spectrally selective mirror has a character of its own because it perceptibly removes something from the incident light; it darkens some reflected colours more than others. A spectrally nonselective mirror darkens all reflected colours equally. If the darkening is noticeable, the mirror is gray. But if it reflects the same proportion of the light that would be reflected by a diffuse reflector which we would call 'white', its light loss is no longer noticeable and, in consequence, its surface qualities are no longer noticeable either. And then we call it 'clear'.

White is thus unique among the colours of the surfaces of objects in that it alone involves another surface property, diffuse reflectance. So McGinn is correct insofar as he claims that high unselective reflectance is not the whole story. But we are now able to see why his fantasy about the Martian is fundamentally misguided. For that story requires a change of appearance like a change of apparent hue, except that the achromatic colours are slotted in among the hues.[1] Now they must be slotted in as a family as if they were of a single hue, since they vary among themselves only in lightness, just as all the pure greens vary among themselves in lightness. We are only entitled to say that the Martian sees achromatically where we see greenly. Now let him see the lightest gray where we see the lightest green, and let the object be transparent. Will what he sees then be a white transparency? The transparency at which he is looking, the one which we see as green, will be spectrally selective, and will necessarily darken some colour (for example the one we call 'red') which is seen through it. This is a physical fact about the transparency, one which will not vary with the psychological constitution of the observer. So what the Martian will see is an achromatic transparency which for some set of wavelengths transmits significantly less light than is available from the original scene. But this means that if he sees the transparency as coloured, he will see it as variably *gray*, and the conceivability of gray transparencies is not at issue.

Finally, consider what happens to his perception of things which we regard as achromatic. He will see them as colour K – but only as long as they diffuse or absorb a noticeable amount of light. Let such an object reflect or transmit all of the light without diffusing it, and that object will have no perceivable surface character of its own. He will regard it not as K, but as *clear*. He may then have a problem about the non-existence of transparent K which is precisely coextensive with our problem about transparent white, and for just the same reasons. Colour sensations have simply dropped out of the equation, and with them any need to appeal to 'a rich system of phenomenological laws'.[2]

Syracuse University,
541 Hall of Languages,
Syracuse, New York 13210, U.S.A.

© C. L. HARDIN 1985

---

[1] It is right at this point that McGinn's tale becomes incoherent, because the achromatic and chromatic families are structurally different and thus cannot be transformed into one another. But I shall not argue this point further here.

[2] I am indebted to Jonathan Westphal's article 'White', which I saw in manuscript, and to Jonathan Bennett's expository suggestions.

VERIDICALITY: MORE ON SEARLE

By ALAN MILLAR

THERE is good reason to distinguish between visual perception (seeing) and visual experience. The distinction is exhibited by the following logical relationships.

(1) Tom has a visual perception of (sees) a car before him now

entails both

(2) Tom has a visual experience of a car before him now

and

(3) There is a car before Tom now.

(2) on the other hand entails neither (1) nor (3). (2) merely states that Tom is having a visual experience with a certain content specified by the phrase 'of a car before him now'. In his book *Intentionality* (Cambridge, 1983), John Searle has argued that the content of visual experiences can be more explicitly specified by means of propositional clauses. As a first step in this direction he would have us replace (2) by

(2′) Tom has a visual experience that (there is a car before him now).[1]

This view has obvious advantages. It serves to highlight important analogies between visual experiences and other intentional states such as beliefs. To know the content of a belief is to know what would have to be the case if it were true. Similarly, to know the content of a visual experience is to know what would have to be the case if it were veridical. The specification of the content of an intentional state is a specification of what Searle calls its satisfaction conditions: those states of affairs which, were they to obtain, would satisfy the state. What satisfies a belief makes it true. What satisfies a visual experience makes it veridical.

A controversial feature of Searle's theory of intentionality is the claim that visual experience is *causally self-referential* (op. cit., pp. 48 ff.). In terms of our example, what this means is that (2′) must be replaced by

(2″) Tom has a visual experience that (there is a car before him now and that there is a car before him now is causing his present visual experience).

The idea is that for Tom's experience to be veridical things must be as they visually seem to him to be and their seeming to him to be that way must depend causally on their being that way. I think Searle is wrong in this and will say why below. First I shall consider a recent criticism of Searle by Gregory McCulloch (ANALYSIS 44.4, October 1984, pp. 203–5).

---

[1] Searle puts the first bracket before the 'that'. By placing it after the 'that' we can make it clear in examples such as (2″) that the second 'that' comes within the scope of the first.

McCulloch focuses on a claim which has the form of

(4) Tom visually perceives that there is a car before him now.

He, like Searle (Searle, op. cit., p. 61, but cf. p. 42), takes (4) to be equivalent to (1). I think this can reasonably be doubted but nothing that I have to say turns on the matter. McCulloch then suggests that Searle conflates two questions:

(i) what are the satisfaction conditions determined by the content-specifying clause in (4) viz. 'there is a car before him now'?
(ii) is (4) true?

If (4) is true Tom must have a visual experience of a car which is caused by there being a car before him now. McCulloch's point, I take it, is that, nevertheless, Tom could have just such an experience even if it were not thus caused. However, it is not immediately obvious why McCulloch thinks that this point constitutes an objection to Searle. Suppose that (4) is in fact true. According to Searle, Tom in that case has a visual experience of the type described by (2″). But he is at pains to rebut the suggestion that one can have an experience with such a content only if it is caused by there being a car before one. (Searle op. cit., pp. 130 ff.) His view is rather that that *if* such an experience were *veridical* (satisfied) then it would have to be caused by there being a car before the person who has it. McCulloch in fact misunderstands Searle's view for he claims (McCulloch op. cit., p. 203) that Searle 'identifies the satisfaction conditions of the content of a visual experience with those conditions that *ensure that a veridical experience is had*' (my emphasis). On the most natural reading of the phrase italicized, the view ascribed to Searle is that the satisfaction conditions which exhibit the content of an experience are those conditions which make the experience veridical, i.e. actual states of affairs. Further evidence that this is the intended reading is provided by McCulloch's later claim (op. cit., p. 204) that '[s]atisfaction conditions are states of affairs in "an independently existing reality"'. Now Searle acknowledges that the expression 'satisfaction conditions' can be used to refer to facts (*actual* states of affairs) which satisfy intentional states (Searle op. cit., p. 13). In this sense the satisfaction condition of a (veridical) visual experience would be a fact which makes it veridical. But that is not the sense which is relevant to the claim that we know the content of an experience when we know its satisfaction conditions. Searle does not hold the view ascribed to him nor does he conflate (i) and (ii).

Let us then take a fresh look at the causal self-referentiality thesis. On the face of it there is no more reason to build a causal ingredient into the contents of visual experiences than there is to build such an ingredient into the contents of beliefs. Suppose that the causal theory of knowledge is true. In that case A knows that p only if A believes that p and his believing that p depends causally on its being the case that p. It does not follow from this that A's belief that p is true only if it depends causally on its being the case that p. All that is required for the belief to be true is that it be the case that p. Suppose now that the causal theory of perception is true. In that case A perceives a ϕ before him only if he has a visual experience of a ϕ before him, and that experience depends causally on there being a ϕ before him. Again it does not follow (even if it were true) that A's experience is *veridical* only if it depends causally on there being a ϕ before him. The causal theory of perception does not by itself provide a reason to build a causal ingredient into the contents of visual experiences and the analogy between visual experiences and beliefs makes it plausible to hold the contrary view that all that is required for A's visual experience of a ϕ to be veridical is that there be a ϕ before him.

However, it is not difficult to see what leads Searle towards the causal self-referentiality thesis. In terms of our example it is the tacit assumption that Tom's visual experience is veridical only if Tom *perceives* a car before him. Indeed I am inclined to think that Searle simply equates having a veridical visual experience of a ϕ before one with visually perceiving a ϕ before one. Granted the causal theory of perception, it follows that for Tom's experience to be veridical, it must causally depend upon there being a car before him. But then since the content of a visual experience is just what would have to be the case if it were veridical, the argument concludes that the content of Tom's experience must include the causal ingredient indicated in (2″). This is not to say that the experience would not have the content it does have unless it (the experience) depended causally on there being a car before Tom, only that the content requires that such a causal relation obtain if the experience is veridical.

The problem with this line of reasoning lies in the initial move, that is, with the tacit assumption that Tom's visual experience is veridical only if he perceives a car before him. My claim to the contrary is that an experience of a ϕ before one can be veridical even if the subject does not perceive a ϕ before him. To make good this claim, an example is required in which (a) a person has a visual experience of a ϕ before him, (b) there is a ϕ before him and (c) the person's experience is not causally dependent on there being a ϕ before him. It must also be shown that (a) and (b) add up to the fact that the experience is veridical in an intelligible and interesting sense of that term.

The examples which naturally suggest themselves involve laser photographs, mirrors or tamperings with the brain. I shall opt for mirrors. Suppose that when Tom has his experience of a car before him, he is actually looking through a doorway at a mirror whose edges he cannot see. Tom is unaware of the presence of the mirror and takes himself to be perceiving a car before him, for the mirror is so angled that it reflects light from a scene involving a car of a certain type against a certain background. Suppose now that behind the mirror in front of Tom there is another scene involving a car. This scene is in appearance the mirror image of the first but in all other respects is like the first. The arrangement results in Tom's having a visual experience which is qualitatively identical to that which he would have were he to perceive the scene which is behind the mirror. Of course he does not perceive the scene behind the mirror for the requisite causal relation does not obtain. Nonetheless it visually seems to Tom that there is a car before him of a certain type against a certain background and at a certain distance, and there is. So conditions (a) to (c) hold. Moreover there is a coherent and interesting notion of veridicality in accordance with which Tom's experience is veridical. On this notion of veridicality, an experience of a ϕ before one is veridical if there is a ϕ before one. What makes it interesting is that it meshes with a notion of content which reflects how things visually seem. That surely is precisely what we want the notion of veridicality to do.

Some people initially find it odd to think of Tom's experience as being veridical in the circumstances envisaged. The sense of oddness derives from two sources. The first is simply the fact that Tom's experience does not depend causally on the scene behind the mirror. But while that is a reason for saying that Tom does not perceive the scene behind the mirror, it is not a reason for denying that his experience is veridical in the sense explained. The second source of the sense of oddness is the fact that Tom's experience is deceptive. This is a more serious point. Its resolution reveals an important feature of the concept of veridicality. Tom's experience is of a car of a certain type etc. but it is also an experience of there being nothing (but air) between Tom and a car before him. Since there is something, a mirror, between Tom and the car before him the experience is deceptive and, surely, not veridical. This is correct. What it shows is that, as one would expect, veridicality is relative to specifications of content. *Qua* experience of a car before him, Tom's experience is veridical. *Qua* experience of there being nothing between Tom and a car before him, it is not. Tom's experience *is* deceptive but this does not undermine the suggested notion of veridicality nor count in favour of Searle's.

It is open to Searle to grant me my use of the term 'veridical' and to claim that nevertheless there is more to the content of a visual experience than how things visually seem. Even so, we still lack a decisive reason for preferring Searle's view to the simpler view which preserves the analogy between visual experience and belief.

There is a line of thought in Searle's book which could be pressed into service at this point, although Searle does not rely on it when he introduces the thesis of the causal self-referentiality of visual experience. Searle thinks that we are 'directly aware of causation in perception' (op. cit., p. 125). What he means is that whenever we have a visual experience, whether veridical or not, we experience it 'as caused by something independent of us'. In support of this, Searle asks us to consider the following thought experiment.

> Suppose we had the capacity to form visual images as vivid as our present visual experiences. Now imagine the difference between forming such an image of the front of one's house as a voluntary action, and actually seeing the front of the house. In each case the purely visual content is equally vivid, so what would account for the difference? The voluntarily formed images we would experience as caused by us, the visual experience of the house we would experience as caused by something independent of us. The difference in the two cases is a difference in the causal content of the two experiences (op. cit., p. 124 footnote).

While this view deserves further examination in the context of the wider concerns of Searle's book, I do not find the argument in the passage quoted to be convincing. It is certainly true that normally when we have a visual experience we instinctively *believe* it to be caused by something in our environment which we take ourselves to perceive. Such beliefs are confirmed or disconfirmed by the ways in which our visual experiences cohere, or fail to cohere, with past and subsequent experience, both visual and in other modalities, in the way that we should expect if the experiences in question were caused by what we think they are caused by. Visual images produced at will would not cohere in the right kind of way. Any difference in 'feel' between such images and visual experiences proper can be explained in terms of memories of preceding experiences and expectations of succeeding experiences. So we still lack a good reason for building a causal ingredient into the contents of visual experiences.[2]

University of Stirling,
Stirling, FK9 4LA

© ALAN MILLAR 1985

---

[2] I am grateful to my colleagues at Stirling and in particular to Andrew Brennan, Murray MacBeath, Tony Pitson and Neil Tennant for comments on ideas contained in this paper.

WHAT IS A DICTATOR?

By LESLIE GREEN

I

THERE is a populous, direct democracy which has exceptionally good luck in the business of making social decisions. Many issues divide naturally into binary choices and, in the rest, there is always one alternative which all voters agree is not the worst. Knowing some social choice theory, they rejoice that this good fortune secures them from much of the manipulability and indeterminacy of social choice in all other democracies. Above all, Arrow's celebrated theorem in inapplicable to their case. The domain of possible preferences being naturally restricted, there are rules for social choice which satisfy the conditions of the Pareto principle, the independence of irrelevant alternatives, and non-dictatorship, while producing a complete and transitive ordering of the alternatives. Democracy seems to flourish until a newspaper reports the true confessions of three miscreants who appear to have succeeded in subverting it. They are:

(1) *Bored*, who finds every vote completely dull because, even with full information and after careful consideration, he always judges each alternative as good as every other.
(2) *Scared*, who so much fears ending up in the minority that he always votes as he predicts the majority will – and who has always been right. And,
(3) *Lucky*, who, by sheer chance (perhaps he is a perfectly representative median voter) turns up on the winning side of every vote.

Now many citizens feel no threat from Bored, Scared or Lucky, but the Department of Social Choice points out that their presence means that, all along, their society has been a dictatorship, for it fails to satisfy Arrow's condition that: 'There is no individual whose preferences are automatically society's preferences independent of the preferences of all other individuals' ([2]: p. 226). The anti-democratic behaviour of the Gang of Three consists precisely in violating this by having preferences which are perfectly correlated with social choices. Whatever anyone else wants, they always get their way.

II

The above tale suggests that, contrary to the insistence of Arrow and almost every other writer, non-dictatorship (*D*) may not be 'the least controversial of all' requirements on a method of social choice

([2]: p. 226). Thinking it uncontroversial has often led to sloppy formulations of the principle and bad arguments for it. W. H. Riker, for example, in an otherwise lucid book, writes that:

> there [should] be no person *i* such that, wherever xP_iy [i.e. *x* is preferred by *i* to *y*] then the social choice is *x* regardless of the opinions of other persons. Since the whole idea of democracy is to avoid such situations, the moral significance of this condition is obvious ([6]: p. 118).

But since the case of Bored, Scared and Lucky is included among these situations, at least some of its violations are inconsequential. Nor does the conceptual appeal to 'the whole idea of democracy' work. Since the satisfaction of *D* is simply treated as part of the definition of democracy, appeal to that definition can have no independent force in the argument.

A similar misplaced conceptual appeal turns up in Arrow, who regards *D* as a restriction necessary for the method of social choice to count as collective ([1]: p. 30). And this argument is reiterated, for example, by A. F. MacKay, who writes, 'If ... one wants an acceptable *collective* choice mechanism, then to have a dictatorial device is irrational. Dictatorship is not collective choice at all' ([5]: p. 22). But the argument from italics is as weak here as it is elsewhere in philosophy. It is crucially indecisive between MacKay's concept of *collective* choice procedures (i.e. choosing by a whole collectivity) and collective *choice* procedures (i.e. choosing for a whole collectivity). Which is the prefered conception depends on a substantive argument about the point of the restriction, and not on the very idea of collective choice.

Some restriction like *D*, can, of course, be justified substantively, by appeal to moral principles of equality or autonomy. And a more careful formulation can rule out cases like that of Bored (see [7]: p. 42). But to deal with Scared and Lucky, we need to narrow the general prohibition on perfect correlations between individual and social preferences. A beginning would be to limit the direction and nature of causality, for we are only concerned about cases where an individual's preferences determine social choices in an unfair way and not simply when they match them perfectly (Scared) or determine them innocuously (Lucky). The overkill of *D* is a consequence of a need to make the formalization tractable by relying only on the calculus of preference relations. But that simplicity is too dearly purchased. We could salvage it by simply imposing *ad hoc* restrictions on preference orderings like those of Bored, Scared and Lucky, but that is to give up the condition of unrestricted domain on which Arrow's theorem depends. And if we are willing to restrict preferences here, then why not elsewhere too? To permit only certain preference orderings as inputs to the social choice mechanism seems as bad as to restrict the number of individuals whose preferences count in determining its output.

III

There are other, independent, difficulties with *D* when social choice theory is understood in the framework of a general behavioural theory of expected utility maximization. As I argued in ANALYSIS 43.1 [4], the economically rational would not participate in any *non*-dictatorial social choice procedure since the costs of sharing power with others are larger than the expected benefits. And even without the factual assumptions on which that argument depends, it still follows that they would regard certain dictatorial procedures as ideally the best ones. They are, after all, egoists.

David Gordon (ANALYSIS 44.2 [3]) thinks this argument relies on an uninterestingly wide conception of a dictator, and that the truth that it does contain is trivial. If only one person were permitted to vote then, Gordon concedes, we would call him a dictator, but if only one turned out to vote we would not. That may be so, but my argument depends on no such ambiguity. The point is rather that the economically rational would be under-motivated by any procedure which allowed others to vote, or at least that they would regard a procedure which did not allow others to vote as being the best. Gordon finds nothing paradoxical about the latter – and neither do I – although he finds it an obvious truth. Here also I agree if he means it obviously follows from the premises. But if he means instead to accept the premises and thus the conclusion that a dictatorial procedure is indeed best, then he may be confused about what that would permit. It is not simply the point Filmer made in his *reductio* of consent theory: 'he were a madman that being by nature free would choose any man but himself to be his own governor'. Perhaps that truth is obvious. However, it is not obvious, but dubious, that one would have to be irrational not to want to govern others as well.

The social choice theoretic concept of a dictator is indeed excessively broad, though not for the reason Gordon gives. It is too broad because it covers unobjectionable cases like those of Bored, Scared and Lucky. Together with my argument that it is inconsistent with certain consequences of the general behavioural theory in which it is embedded, this should be enough to make us question its formulation and status. Because Arrow's theorem is valid, philosophical scrutiny must centre on its premises. It is time to add non-dictatorship to the list of those in dispute.

Lincoln College, Oxford OX1 3DR

© LESLIE GREEN 1985

REFERENCES

[1] Kenneth J. Arrow, *Social Choice and Individual Values* (New Haven: Yale University Press, 1963).

[2] Kenneth J. Arrow, 'Values and Collective Decision-Making', in P. Laslett and W. G. Runciman (eds), *Philosophy, Politics, and Society*, third series (Oxford: Blackwell, 1967).

[3] David Gordon, 'Green on Dictators and Democracies', *Analysis* 44.2, March 1984, pp. 95–6.

[4] Leslie Green, 'Dictators and Democracies', *Analysis* 43.1, January 1983, pp. 58–9.

[5] Alfred F. MacKay, *Arrow's Theorem: The Paradox of Social Choice* (New Haven: Yale University Press, 1980).

[6] William H. Riker, *Liberalism Against Populism* (San Francisco: W. H. Freeman, 1982).

[7] Amartya K. Sen, *Collective Choice and Social Welfare* (Amsterdam: North-Holland, 1979).

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 63, No. 2; June 1985
EDITOR: BRIAN ELLIS

ARTICLES

Graham Nerlich and Andrew Westwell-Roper	Ontology Can Be About: A Spacetime Example
Frederick W. Kroon	Theoretical Terms and the Causal View of Reference
Eric Russert Kraemer	Beliefs, Dispositions and Demonstratives
Gregory Mellema	Shared Responsibility and Ethical Dilutionism
Richard Sylvan	Prospects for Regional Philosophies in Australia

DISCUSSION

James Stephens and Lilly-Marlene Russow	Brains In Vats and the Internalist Perspective
Norman Melchert	Why Constructive Empiricism Collapses into Scientific Realism

CRITICAL NOTICE

Michael Devitt	*The Varieties of Reference* by Gareth Evans

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$17 (US$18 or Stg£10); Institutions A$34, (US$36 or Stg£18). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; Andrew Harrison, R. H. Holland, J. H. McDowell, D. H. Mellor, M. A. Stewart, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices are as follows:

	U.K.	Overseas	U.S.A.	Canada
Institutions	£11.00	£13.50	$33.00	$39.50
Individuals	£7.45	£9.25	$21.00	$25.00

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

[illegible] 1985

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

U.S. MAILING. Second class postage paid at New York. Postmaster: Send address corrections to: Analysis: c/o Expediters of the Printed Word Ltd, 515 Madison Avenue, New York, NY10022 (U.S. Mailing Agent).

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

ol. 45 No. 3 (New Series No. 207) June 1985

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

A problem about ambiguity in truth-theoretical semantics — L. JONATHAN COHEN
An argument for the vagueness of vague — ROY A. SORENSEN
The title of this paper is ‘quotation’ — LAURENCE GOLDSTEIN
Indeterminate people — HUGH S. CHANDLER
Are mental events in space-time? — P. F. GIBBINS
Einstein, Gibbins and the unity of time — MICHAEL LOCKWOOD
A note on temporal parts — H. W. NOONAN
Swain on the basing relation — JONATHAN L. KVANVIG
Short on Teleology — A. OLDING
On two arguments for compatibilism — PETER VAN INWAGEN
Causal deviancy and multiple intentions: a reply to James Montmarquet — JOHN BISHOP
Lyons and tigers — PAUL SIMPSON
Hare on moral conflicts — IGOR PRIMORAC

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

The Philosophical Quarterly

Vol. 35 No. 139 April 1985

Edited in the University of St. Andrews by an Editorial Board
Chairman of the Board of Editors: LESLIE STEVENSON
Executive Editor: JOHN HALL *Reviews Editor*: PETER CLARK

Hume, Animals and the Objectivity of Morals	*Stephen Clark*
Transcendental Idealism in Wittgenstein, and Theories of Meaning	*A.W. Moore*
Ethical Consistency Principles	*Harry F. Gensler*
Anti-Realism and Recognitional Capacities	*A.F. Clark*

DISCUSSIONS

Wollheim and the Institutional Theory of Art	*Graham McFee*
Quasi-Realism and Mind-Dependence	*Stig Alstrup Rasmussen*
Nozick on Inferential Knowledge	*James G. Mazoué*
Transmission for Knowledge Not Established	*Anthony Brueckner*
Response to Mazoué and Brueckner	*Graeme Forbes*

Subscriptions for 1985

Annual Subscription	UK	Overseas	USA	Canada
Institutions	£26.00	£32.50	$57.50	$71.50
Individuals	£10.25	£12.50	$26.00	$30.00

Orders, remittances and enquiries should be addressed to: Journals Department,
Basil Blackwell Ltd., 108, Cowley Road, Oxford OX4 1JF, England.

Journal for the Theory of Social Behaviour

General Editors: C.W. Smith, P. Collett,
and S. Moscovici

Volume 15 no. 2 July 1985

CONTENTS

Explaining Masochism	VIRGINIA L. WARREN
Power, Structure and Agency	DEREK LAYDER
Toward a Social Theory of Ignorance	MICHAEL SMITHSON
Life as Fiction	KEVIN MURRAY
Illustrating Behavioral Principles with Examples from Demography: The Causal Analysis of Differences in Fertility	LINCOLN H. DAY

Published March, July and November
1985 subscriptions:
Institutions £32.75 (inland) £40.95 (overseas) $81.00 (USA) $102.50 (Canada)
Individuals £14.50 (inland) £18.15 (overseas) $29.75 (USA) $39.95 (Canada)
Single copies are available, please write for prices.

Orders to: JOURNALS DEPARTMENT, **BASIL BLACKWELL**,
108 Cowley Road, Oxford OX4 1JF, England.

Invitation to Philosophy

MARTIN HOLLIS

'I read it with pleasure and consider it a good introduction to philosophy.' A.J. Ayer
'Lively, attractive and full of arguments. It offers an invitation to philosophy that many people will want to accept.' Bernard Williams
200 pages, hardback **£14.50** (0 631 14225 8)
paperback **£4.95** (0 631 14226 6)

John Locke: An Introduction

JOHN W. YOLTON

It has been a feature of recent studies of Locke that certain aspects of his thought have been ignored in favour of others. John Yolton's definitive introduction corrects this bias and explores the full spectrum of Locke's work. The result is a significant contribution to Locke studies.
176 pages, hardback **£17.50** (0 631 13376 3)
paperback **£5.95** (0 631 14062 X)

Liberty and Equality

Edited by ELLEN FRANKEL PAUL
With Fred D. Miller and Jeffrey Paul

This book includes articles by R.M. Hare, Jan Narveson, Michael Levin, Werner Dannhauser, William Kristol, Maurice Cranston, Allen Buchanan, James M. Buchanan and Loren E. Lomasky, J.R. Lucas and David Braybrooke.
284 pages, paperback **£7.50** (0 631 13718 1)

Utility and Rights

Edited by R.G. FREY

These essays confront one of the most fundamental questions of contemporary philosophy: the foundation of rights and the primacy of the demands of justice over the claims of welfare. The contributors are L.W. Sumner, Joseph Raz, R.G. Frey, J.L. Mackie, R.M. Hare, H.J. McCloskey, James Griffin, Jan Narveson, Alan Ryan, Rolf Sartorius and Charles Fried.
192 pages, **£19.50** (0 631 13713 0)

Freud and Human Nature

ILHAM DILMAN

'His boldness makes him constantly ask excellent questions . . . a helpful and interesting book.' *British Book News*
This is a thorough, critical discussion of the various issues raised by Freud's ideas of sexual and moral development.
216 pages, hardback **£15.00** (0 631 13373 9)
paperback **£5.95** (0 631 14227 4)

Basil Blackwell

In Defence of Animals

Edited by PETER SINGER

In this book Peter Singer brings together writers and activists in the animal liberation movement, with contributions from others whose work is directly relevant to its aims. He also puts the movement into historical perspective as an extension of basic moral principles.
240 pages, hardback **£15.00** (0 631 13208 2)
paperback **£4.95** (0 631 13432 8)

A Question of Life

MARY WARNOCK

This new edition of the Warnock Report on human fertilization and embryology includes two new chapters: in one Mary Warnock explains the background to the Report, examining the complex, conflicting and heartfelt views involved; in a new conclusion she widens the issue by considering the advisability of expecting a small group to provide the ethical basis for legislation of such great social significance.
128 pages, paperback **£4.95** (0 631 13257 6)

Modes of Occurrence

BARRY TAYLOR

This book provides a novel theory of abverbial structure within the context of a rigorous account of the nature of events. It is a comprehensive and elegant account with a full appreciation of semantic and metaphysical effects, and its original twists and attention to complexity of detail ensure that it will be regarded as an important and innovative work.
160 pages, **£12.50** (0 631 14026 3)

A Dictionary of Marxist Thought

Edited by TOM BOTTOMORE

Editorial Board: Laurence Harris, V.G. Kiernan, Ralph Miliband

'Indispensable . . . it is of enormous scholastic value and will I am sure become even more valuable . . .' *New Statesman*
'The entries cover a very wide range . . . clear, accurate and concise.'
Times Higher Education Supplement
602 pages, paperback **£8.50** (0 631 14236 3)

Spheres of Justice

A Defence of Pluralism and Equality

MICHAEL WALZER

'Walzer's discussions display a rare sensitivity to the uniqueness of each problem: they are stimulating and imaginative in a way which lives up to the high expectations aroused by his earlier writings in political philosophy.' Jeremy Waldron, *Times Literary Supplement*
'Remarkable achievement . . .' Steven Lukes, *The Observer*
364 pages, hardback **£19.50** (0 85520 683 7)
paperback **£7.95** (0 631 14063 8)

Basil Blackwell

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 63, No. 3, September 1985
EDITOR: BRIAN ELLIS

ARTICLES

John Watkins	On Stove's Book, by a Fifth 'Irrationalist'
Ferdinand Schoeman	Aristotle on the Good of Friendship
Gregory Currie	The Analysis of Thoughts
Denis Robinson	Can Amoebae Divide Without Multiplying
Jan Crosthwaite	The Meaning of Metaphors
Quentin Smith	The Anthropic Principle and Many-Worlds Cosmologies

DISCUSSION

Peter van Inwagen	Compatibilistic Reflections
Richard Feldman	Schmitt on Reliability, Objectivity, and Justification

CRITICAL NOTICE

Ross Phillips	*Rationality and Relativism* by Martin Hollis and Steven Lukes

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$17 (US$18 or Stg£10); Institutions A$34, (US$36 or Stg£18). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

The Journal of Philosophy

Subscriptions at $15.00/year; $12.00 to students
$20.00/year to libraries

Students' price extended to retired and unemployed philosophers. Published bimonthly 1904 to 1976, monthly thereafter. Complete volumes and all separate issues available back to January 7, 1904 (volume 1, number 1); prices as follows:

BACK VOLUMES (from the first to the penultimate year)
32-page issues, $1.50; 64-page, $3.00; special issues, $4.00. Volumes (unbound, volume indexes included), $30.00 each.

CURRENT VOLUMES (this year and last)
64-page issues, $2.00; special issues, 96-page or over, $3.00. Volumes (unbound, volume indexes included), $25.00 each.

ALSO AVAILABLE
Cumulative Fifty-year Index, 1904–1953; articles classified by subject and author; 452 p.; cloth, $12.00. Ten-year Supplement, 1954–1963: 98 p.; cloth $3.00, paper $2.00.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

ANALYSIS 45.3 JUNE 1985

A PROBLEM ABOUT AMBIGUITY IN TRUTH-THEORETICAL SEMANTICS

By L. JONATHAN COHEN

IF one aims, as Davidson does, to build a semantics for a natural language on the foundation of Convention T, how far does one have to move away from the Tarskian paradigm in order to cope with the existence of ambiguity in many sentences of natural language? I shall argue that one needs to move a good deal further than Davidson supposes.

Obviously syntactic ambiguity, as in [1]

Someone heard the shooting of the hunters,

creates no special difficulties here. An adequate grammatical syntax will generate unambiguous structural descriptions, and under such a description a sentence like this can, as appropriate, either be said to be true if and only if someone heard the hunters shooting or be said to be true if and only if someone heard the hunters being shot. What precise form the postulates will need to have, in order that such T-sentences may be derivable, would obviously take quite a bit of working out, even where object- and meta-language are the same. But the extra dimension of complexity introduced into that problem by the ambiguity would call for a grammatical rather than a logical or philosophical resolution.

Verbal ambiguity, however, creates problems that are not so easy for philosophers to brush aside. Davidson's first attempt (in 'Truth and Meaning', *Synthese* 17, 1967, p. 316) to meet this problem was to argue that, if a verbally ambiguous sentence can preserve its ambiguity when translated into the metalanguage, 'a truth definition will not tell us any lies'. No doubt he had in mind such sentences as

(1) 'The box was in the pen' is true for an English-speaker at time *t* if and only if the box was in the pen before *t*.

And he claimed that in this respect ambiguity was just like vagueness or even like the property of being unsuitable for incorporation in a serious science. Where object- and metalanguage are the same, he said, 'all *these* problems will be translated without loss or gain into the metalanguage'.

---

[1] For a defence of the existence of syntactic ambiguity here see A. M. Zwicky and J. M. Sadock, 'Ambiguity Tests and How to Fail Them', in J. P. Kimball (ed.), *Syntax and Semantics* Vol. 4, New York, Academic Press 1975, p. 13.

Unfortunately vagueness and ambiguity cannot be treated alike in this regard. Consider for example the difference between

(2) George's time of arrival was fiveish in the afternoon, and
(3) George arrived at 4.57 pm.

Obviously (2) is relatively vague, at least when compared with (3). And it would be quite appropriate to preserve this vagueness on the right-hand side of the relevant T-sentence. Sentences with relatively vague meanings are often needed for our self-expressive or communicative purposes, and statements of their truth-conditions must allow a correspondingly indeterminate spread of possibilities. So the sentence (2) is true if and only if George's time of arrival was fiveish in the afternoon, and to know that this is so might plausibly be said (at any rate from a realist point of view) to be part at least of what it is to know the meaning of (2). But the expression 'pen' is not a particularly vague one, at least in respect of the fact that it may signify either an instrument for writing with ink or an enclosure for ungulates or infants. Nor can knowledge of this ambiguity, which every competent English-speaker has, be usefully expressed in ambiguous terms, as in (1). An ambiguous description of some knowledge that a person has is not a scientifically satisfactory description of that knowledge, even if the knowledge in question is about an ambiguity. To assert such a description would be to make a pun, and play on words is scarcely a serious contribution to the investigation of linguistic competence.

It must have been for some such reason as this that three years later (in 'Semantics for Natural Language', in B. Visentini *et al.*, *Linguaggi nella Società e nella tecnica*, Milan, Edizioni di Communità, 1970, p. 182), without referring at all to his earlier account, Davidson proposed a quite different treatment of verbal ambiguity. Davidson argued that a sentence like 'The box was in the pen' created no insoluble problems for a theory of truth because it might be accommodated within a T-sentence like

(4) 'The box was in the pen' is true for an English-speaker x at time t if and only if either the box was in the playpen before t and the circumstances surrounding x at t meet condition c, or the box was in the writing-pen before t and the circumstances surrounding x at t meet condition c'.

Whether such a theory would then be acceptable, Davidson claimed, would depend on what we were prepared to allow in the description of the relevant conditions c and c'.

But some other issues also enter into the acceptability of the ensuing theory. Whatever form they take, conditions c and c' cannot jointly exhaust the possibilities. Some circumstances must be possible in which neither c nor c' prevails. Otherwise one would not be able to appreciate the inherent ambiguity of the sentence so well. And in such circumstances the predicate 'is a pen' must fail of satisfaction. Davidson's assignment of truth-conditions to verbally ambiguous sentences thus resembles, in one respect, Russell's assignment of truth-conditions to definite descriptions. Where none of the conditions disjunctively presupposed by the utterance of a certain verbally ambiguous predicate are actually present, the predication is false.

Perhaps it will be said that in constructing a semantics for natural language one need not be too worried about undisambiguated assertions of ambiguous sentences. Such assertions are deviant by the standards of ordinary conversational norms, so why not accept the convenience of being able to reckon them all false?

The answer to this objection is twofold. Where the ambiguity is a deliberate pun the point of the pun would be lost if the assertion had to be regarded as false. And where the ambiguity of the assertion is due to carelessness in matching sentences to circumstances appropriate for their utterance the author of the assertion is being convicted of a crime that he did not commit (viz. the crime of asserting what is not the case) and not convicted of the crime that he did commit (viz. speaking carelessly). Yet to prevent such anomalies it looks as if the semantics would have to admit the possibility of truth-value gaps.

To avoid having to make this admission it may be tempting to claim that verbal ambiguity is itself a rare, deviant and therefore quite unimportant occurrence in the sentences of a natural language. Homonyms like '(play)pen' and '(writing)pen' or '(legal) case' and '(suit)case' are not only very uncommon, it may be said, but also do nothing to improve a language's flexibility or intelligibility in everyday conversation. They are just accidents of etymological history. A semantics that was adequate for the rest of natural language would then be adequate enough.

But it is plainly wrong to suppose that the problem of verbal ambiguity in natural language can be satisfactorily resolved by tracing all such ambiguity to the existence of admitted homonyms. The fact is that verbal polysemy, i.e. a word's having more than one sense, so far from being a rare and pathological phenomenon like homonymy, is the rule rather than the exception in natural language. Scan the pages of any English-English, English-French, French-English or French-French dictionary and the evidence lies in front of you. The words *not* assigned more than one standard sense are the odd ones: recent foreign imports, scientific coinings, thieves' argot, contemporary slang, and so on. The rest are polysemes. This is equally true of adjectives, as in 'healthy person' and 'healthy place', of nouns, as in 'table of mahogany' and 'table of figures', and of verbs, as in 'drops a bomb', 'drops a hint', 'drops a friend' or 'drops his voice'. Even prepositions have several familiar meanings: if George is 'in' the pool, the pool must be larger than George, but if George is 'in' trouble or 'in' the mood to go swimming, it does not follow that George's trouble or mood is comparable for size with George himself. Maybe there is a sense in which most of these phrases can be described as 'dead metaphor'. But if so most of the language we ordinarily use (including the phrase 'dead metaphor' itself) is dead metaphor. A semantics that ignores such polysemy thereby abandons any claim to be a semantics for natural language, because it ignores one of the main features —a feature comparable with indexicality — that differentiates a natural language from a Tarskian artificial one.

So the problem of verbal ambiguity cannot be swept under the carpet as the outcome of some pathological abnormality. Rather, it is an enormous convenience that the same word can often be uttered in one or other of several different though related senses. Instead of having to learn a very much larger number of words, each with fixed and context-independent meanings, we can learn a relatively small number of words with variable meanings and then exploit their verbal or situational contexts of utterance in order to disambiguate their actual occurrences. The economy of resources that is thus achieved by natural language is obviously comparable with the economies that are achieved by variability of reference in the case of proper nouns and other indexical expressions as an alternative to the fixed references available with the use of co-ordinate systems, individual constants, etc.

Moreover, if all ambiguous occurrences of such polysemes are to be dealt with as in (4), the question arises whether the disjunctions on the right-hand side of the ensuing T-sentences are always capable of being written down. Not only does the language need to be enormously rich in synonyms if this is to be possible, since so many potential disambiguations have to be expressible. But also the disjunctions have to be finite in length.

Can such finiteness be guaranteed? Consider the way in which a word that names a kind of object, situation or substance is also available to denote a two- or three-dimensional representation of an instance of this kind, as in

(5) Four lions surround the column.

Whether this usage exists in all natural languages, I do not know. But it is familiar in modern European languages, and its existence in ancient Greek was remarked by both Plato (*Republic* 597B) and Aristotle (*Categories*, ad init.). And the usage does not have to be learned separately for each noun denoting a species of animal, plant, artefact, event, etc. There may well be at least one such English noun that has never yet been uttered in a representational sense. Indeed that is presumably why the big Oxford dictionary (*The Compact Edition of the Oxford English Dictionary*, Oxford, Clarendon Press, 1984) does not list either of the words 'rabbit' or 'zebra' in this sense though it does list 'lion' and 'tiger'. But in relation to any such word we do not have to await an utterance of the relevant kind before being willing to accept that the representational sense is available in the language. In suitably disambiguating circumstances that sense is immediately understood. Nor is this limited to first-order representations. The word 'face', say, can not only come to mean – in appropriate circumstances – a representation of a face but also a representation of a representation of a face, and so on, as becomes apparent immediately one considers what one could say about a picture of a picture gallery in which one of the pictures is also a picture of a picture gallery. The semantic mechanism here is recursive, in that if a word can signify x it can also signify a representation of x. So it is arguable that (5) is, in principle, infinitely ambiguous.

Or consider how the meaning of the word 'drop' changes from 'drop studying' in

(6) Most students here drop geography in their final year

to 'drop attending' in

(7) Most students here drop geography lectures in their final year,

and then to 'drop executing' in

(8) Most students here drop geography lectures reading-assignments in their final year,

and then to 'drop paying' in

(9) Most students here drop geography lectures reading-assignments library-fees in their final year.

If sentences can be of any finite length, there seems no limit to the variety of meanings that surrounding phrases can impose on words like 'drop', 'make', 'start', 'establish', etc. So

(10) Most students here drop them in their final year

looks infinitely ambiguous.

Certainly the variations in the meaning of 'drop' in (6)–(9) cannot be put down to vagueness or generality rather than ambiguity. They cannot be regarded as meaning different species of a genus rather than as being different meanings of a word. This is evident from such sentences as

(11) John has dropped his geography lectures and Peter has done so also.

Here 'dropped' may imply 'dropped giving' in both cases, or it may imply 'dropped attending' in both cases. But it cannot, except as a rather poor joke, be used to refer to one kind of thing in John's case and to another in Peter's, though this would be quite acceptable in

(12) John has lost a pet and Peter has done so also,

where John may have lost a cat, say, and Peter a dog.

So it looks as though any T-sentence that stated the truth-conditions for (5) or (10) on the model of (4) would have to be infinitely long and the relevant satisfaction-conditions would neither be learnable nor statable in a finite period of time. Of course, it is easy enough to capture these infinities informally, by treating it as a semantic principle that, if a word can signify x it can also signify a representation of x, or that, if an agent does NP if and only if he or she VP's NP, then the agent drops NP if and only if he or she drops VP-ing NP. But, if a substantial number of such principles have to be grafted on to a truth-theoretical semantics in order to make sufficient provision for verbal ambiguity, the original felicity of the Tarskian paradigm is further compromised.[2]

The Queen's College,
Oxford
©L. JONATHAN COHEN

---

[2] An earlier version of this article formed part of a paper presented at the May, 1984 conference on Davidson's work in New Brunswick, N.J. I am grateful for comments by several members of the audience on that occasion and for other comments by Jonathan Adler and Alan Weir.

AN ARGUMENT FOR THE VAGUENESS OF 'VAGUE'

By ROY A. SORENSEN

ALTHOUGH some commentators on vagueness have claimed 'vague' is vague, none has presented a clear, intuitive argument for the thesis. I think such an argument can be provided with the help of a certain sequence of disjunctive predicates. After describing these predicates and presenting the argument, I will show how the vagueness of 'vague' creates difficulties for those who believe vague predicates are incoherent and those who wish to deal with the problems of vagueness by a demarcation between vague and non-vague predicates.

Here is one version of the sorites paradox:

(1) 0 is a small number.
(2) If n is a small number, then n + 1 is a small number.
So (3) One billion is a small number.

As far as I know, all commentators agree that the defectiveness of this mathematical induction is somehow due to the vagueness of 'small'. Although there is considerable disagreement over the nature of the defectiveness and the exact nature of vagueness, there is general agreement that predicates which possess borderline cases are vague predicates. Commentators provide conflicting definitions of 'borderline case'. Some define it behaviourally, some epistemically, still others define borderline cases in terms of truth value gaps. My use of the notion of borderline case is intended to be compatible with all of the proposed definitions.

Now consider the sequence of numerical predicates '1-small', '2-small', '3-small' etc. The n-th predicate on the list is defined as applying to only those integers that are either small or less than n. These predicates can be used to construct a sorites paradox for the predicate 'vague'.

(1) '1-small' is vague.
(2) If 'n-small' is vague, then 'n + 1-small' is vague.
So (3) 'One-billion-small' is vague.

The predicate '1-small' is just as vague as 'small' because both predicates clearly apply to 0 and apply precisely in the same way to all the other integers. The same holds for '2-small' and '3-small', but eventually we reach predicates where the 'less than n' clause eliminates some borderline cases. Once we reach predicates in which all borderline cases are eliminated, we can tell that 'vague' no longer applies. For instance, it is a clear cut issue whether the predicate 'one-billion-small' applies to a given integer or not: since no integer greater than a billion is small, this predicate is as definite as the predicate 'less than one billion'. But it is unclear as to where along the sequence the predicates with borderline cases end and the ones without borderline cases begin. In short, 'vague' is vague.

The vagueness of 'vague' suggests that accounts of vagueness which describe vague predicates as incoherent are necessarily false. Such an account can be true only if the terms it *uses* to describe vague predicates are coherent. Since they use 'vague' in their thesis that all vague predicates are incoherent, their account can be true only if their account uses an incoherent term.

Bertil Rolf, W. V. Quine, and Michael Dummett are recent examples of philosophers who are sympathetic to the incoherence thesis.[1] The most straightforward acceptance of the incoherence thesis appears in the work of Peter Unger and Samuel Wheeler.[2] Unger and Wheeler believe that the sorites paradox shows that vague predicates are incoherent. They point out that the predicates we have for ordinary things are vague predicates. Since incoherent predicates lack extensions, they draw the sceptical conclusion that ordinary things (chairs, flowers, people etc.) do not exist. If Unger and Wheeler were to concede that 'vague' is vague, they would be forced to concede that 'vague' lacks an extension. But then they would have to conclude that nothing is vague. This conflicts with their assertion that there are vague predicates, viz. our predicates for ordinary things. In other words, the following set of propositions are jointly inconsistent.

---

[1] See Rolf's chapter on the sorites in his *Topics in Vagueness* for his rejection of the consistency of vague predicates. Quine's 'acceptance of the paradox and the consequent inconsistency of the pristine use of ostensively acquired terms' is indicated in the sixth and seventh paragraphs of 'What Price Bivalence?' *Journal of Philosophy* vol. LXVIII no. 2, February 1981, pp. 90–5. Dummett's conclusion that a wide class of observational predicates are incoherent appears on p. 319 of 'Wang's Paradox', *Synthese* vol. 30 no. 3/4, April/May 1975, pp. 301–24.

[2] Unger and Wheeler have presented their views in a number of articles. Two representative articles are Unger's 'There Are No Ordinary Things' and Wheeler's 'On That Which Is Not', both appearing in *Synthese* vol. 41 no. 2, June 1979.

(i) There are vague predicates.
(ii) Vague predicates are incoherent.
(iii) Incoherent predicates lack extensions.
(iv) 'Vague' is a vague predicate.

Indeed, since (iv) implies (i), the subset {(ii), (iii), (iv)} is also a set of jointly inconsistent propositions.

Other commentators have complained that proponents of the thesis 'All vague predicates are incoherent' have used vague predicates in their arguments supporting the incoherence thesis. But it might be replied that the use of those vague predicates was an inessential use. However, proponents of the incoherency thesis must make essential use of a vague predicate since 'vague' or an equivalent term must appear in the statement of their thesis. For suppose they restate their thesis with another predicate such as 'nonprecise' or 'borderline-case-inducing predicate'. The substitution will have to be synonymous with 'vague' in order to express the same thesis. Since all synonyms of vague predicates are themselves vague, the thesis will remain a self-refuting thesis.

The vagueness of 'vague' also poses a problem for those who hope to avoid the difficulties of vagueness through a demarcation between vague and nonvague predicates. All complements of vague predicates are themselves vague. Hence, 'nonvague' is vague. Therefore there is no way to determine the exact extension of 'nonvague predicate'. Consider the Frege/Russell principle that logic only applies to nonvague predicates. Does logic apply to this principle? Since it contains a vague predicate, one is tempted to *deduce* from the principle that logic does not apply to the principle. But the deduction is possible only if logic applies to the principle. This self-referential difficulty is a symptom of the fact the principle radically restricts the scope of logic. Moreover, acceptance of the principle would leave the scope of logic indeterminate. For the vagueness of 'vague' ensures the impossibility of exactly determining which predicates are suited for logical evaluation. Matters are not helped by further restricting logic to clearly nonvague predicates because 'clearly nonvague' is vague. For in addition to borderline cases, vague predicates have borderline borderline cases and thus borderline clear cases. Thus the vagueness of 'vague' presents a dilemma for those who wish to restrict logic to nonvague terms.

The statement of the restriction must contain a vague term. If logic applies to the statement, the statement is incorrect. If logic does not apply to the statement, then the 'restriction' is without force; for it has no implications as to what is ruled in or ruled out. Since a restriction must rule something out, the 'restriction' would not be a genuine restriction.

In general, it appears that the vagueness of 'vague' ensures that the worrisome features of vague predicates will also be features of our attempts to describe those predicates. Thus dissatisfaction with vagueness seems to commit us to dissatisfaction with out attempts to express that dissatisfaction.[3]

University of Delaware,
Newark, Delaware 19716, U.S.A.

© ROY A. SORENSEN 1985

---

[3] In writing this paper, I have benefited from the encouragement and remarks of Peter Unger as well as from several important corrections and helpful suggestions from an anonymous referee.

THE TITLE OF THIS PAPER IS 'QUOTATION'

By LAURENCE GOLDSTEIN

QUOTATION may appear too trivial a topic to trouble philosophers. Yet an adequate theory of quotation has proved elusive, and catching this minnow should provide us with bait for bigger fish to fry (Davidson, [3] p. 79, mentions three such areas of concern – sentences about propositional attitudes, explicit performatives and picture theories of reference. A fourth is the Liar paradox, as I indicate in [8]). The classical theory, promoted by Tarski ([12], p. 159), Quine ([9], p. 26) and others, says that a quotation-expression (e.g. a word together with surrounding quotation marks) is a semantically unstructured singular term. That theory is clearly false. One could not see that the statement '"The successor of two" denotes three' is true unless one attended to the composition of the phrase within the inner quotes. Further objections to the classical theory are marshalled by Searle [10], [11], Davidson [3] and others.

Davidson offers a *demonstrative* theory of quotation according to which what appears inside quotation marks is an *inscription*, and the quotation marks may be read 'the expression with the shape pictured here' ([3], p. 90). What Davidson says here of quotation marks he would say of any other quoting device. There are, as he points out, many such devices. In speaking, we sometimes use phrases like 'and I quote', or signal quotation by finger-dancing gestures; one might add that frequently we need recourse to nothing so overt, and rely just on subtle aids such as a queer tone of voice or a slight hiatus.

In [7], I criticized Davidson's theory on the grounds that, even though it is helpful to regard quotation marks as pointing devices, the inscription being pointed to is frequently not being pointed to as a sample of a *shape*. Two examples I cited were Elvis Presley's 'Baby, don't say "don't"' and Gottlob Frege's 'The concept "horse" is not a horse'. If, in these cases, the quotation marks are read in the way that Davidson recommends, the result is absurd. This is hardly surprising, since neither the king of modern rock nor the father of modern logic was referring, in these instances, to (expressions with any) particular (typographical or phonetic) *shapes*. A further, perhaps more contentious, example: the quotation marks occurring in the title of this paper are being used to refer not to a shape, but to a *title* identical not to the title of this paper but to Davidson's [3]. In a translation of that paper into German, the *same title* would be written as 'Die Ausführung'. (Conversely, we don't say that 'Jam dies', the well known Latin poem about daybreak, has the same title as that of a certain affectionate obituary of the Indian mathematician Ramanujam.)

Davidson goes wrong in the very first sentence of his paper. He writes: 'Quotation is a device used to refer to typographical or phonetic shapes by exhibiting samples, that is, inscriptions or utterances that have those shapes' ([3], p. 79. This is incorrect because in *indirect* quotation we don't refer to shapes by exhibiting *those* shapes. In English, the most common indirect quotation functor is 'that'. When someone says 'MacArthur said that *he would return*', he (the speaker) is referring to the utterance 'I shall return' by exhibiting words of a quite different shape.

As Searle urges against a theory Davidson defended in [2], in a faithful indirect report a speaker has to repeat not another persons *words*, but *what that person said* when he uttered those words, viz. his *proposition*. I think that Searle's account of indirect speech, in [11] pp. 180-93, is dead right, and would add to it just one ontological point: to accept that, when quoting, a reporter may be pointing to a *proposition* need cause no embarrassment even to the most hard-nosed physicalist (see my [4], [5] sect. 3, [6] sect. 6). The reporter is just demonstrating, or drawing attention to some *aspect* (here the propositional aspect) of the reportee's speech act while *ignoring* certain features of that act. A faithful indirect reporter can neglect the exact words used by the reportee; indeed, his report can be wholly couched in a language different from the one that the reportee used.

There are occasions when it is not the proposition expressed by a person that is reported, but *what that person means*. Again, no ontological anxieties should attach to the idea that the reporter concentrates on, i.e. points to, only this aspect of the speaker's act while neglecting all its features that do not contribute to what the speaker means. Quotation marks are frequently employed in such reports. For example, 'What he meant was "Let us drink a toast to our dear Queen"' reports what the speaker meant by pointing to a token that *means the same*. (What the speaker actually said was 'Let us drink a toast to our queer Dean'.)

The theory of quotation advocated here and in [7] is, like Davidson's, a *demonstrative* theory, but I diverge from Davidson in maintaining that quotation devices may be used to indicate not just typographical or phonetic shapes, but also such other aspects as *concepts*, *meanings*, *phemes*, *rhemes* and *propositions*. I claim that this is a simple theory that both accounts for all the data and helps to solve a number of puzzles, as any good theory should.

D. K. Buckner [1] criticizes the demonstrative account, but his criticisms are founded on misunderstanding. Buckner thinks that part of what the demonstrative theory says is that we do not read inside quotation marks. Now, any theory that incorporated *that* claim would be ludicrous – how could any sentence containing a quotation expression be understood if the quoted material were not read? The classical theory calls for a *minimal reading* of the quoted material, and regards that material as a semantically undifferentiated whole. But, as we have seen, the demonstrative theory insists, against this, that the understanding of many sentences containing quotation-expressions can be achieved only if we attend to the semantical structure of what lies within the quotation marks. This is particularly clear in cases of 'mixed' use and mention. An example of such, given by Davidson ([3], p. 81) is

> Quine says that quotation '. . . has a certain anomalous feature'

where the mentioned words are obviously also *used*. They are used to say something, and so must be read to learn what is said.

What could have led Buckner to his strange misconception about the demonstrative theory? In [7], p. 4, I say that the quoted token does not occur as a grammatical part of the sentence, and at one point in [3] Davidson says that the quoted material is no part, semantically, of the quoting sentence (though he later (p. 91) qualifies this claim). Buckner takes these remarks to imply that, on the demonstrative account, quotation-expressions can be deleted from sentences with impunity. By performing such an operation on a sentence of Davidson's, Buckner produces 'But if I say or, my hands, or the fish, do not become a part of the language' – a piece of nonsense, as he is eager to point out. This constitutes, so he believes, a *reductio ad absurdum* of the demonstrative theory of quotation.

Buckner's argument is quite specious. He is led into error by failing to distinguish between the *quoted token*, i.e. the material *within* the quotation marks and the *quotation-expression*, which is that material *together with* the quotation marks. Davidson notes

([3], p. 82) that this confusion is common among many who write about quotation, so Buckner should have been forewarned. When Davidson says 'We could easily enough remove the quoted material from the heart of the sentence' he is not in the least suggesting that the quotation-expression be excised, but that the quoted material be removed, that is, *relocated* (not *obliterated*). The point is, perhaps, more clearly made in a sentence of mine that Buckner cites: 'One can think of ... quotation marks as arrows pointing to a sample which could be remote from the sentence'. The purpose of this thought-experiment is to make plausible the idea that quoted words are on exhibition in sentences. This special kind of occurrence could be indicated by enclosing those words in a frame, printing them in an exotic font or, as I suggested, placing them at the end of the sentence, or elsewhere on the page, and using arrows or the indexical 'this:' to point to them.

Far from suggesting that quotation-expressions are dispensable, the demonstrative theory insists that the purpose of quotation marks is to draw attention to the quoted material, for that material serves as a sample. In Davidson's version, it serves as a sample of a shape; in my version contextual factors determine what it is a sample of. The demonstrative theory, but, so far as I know, no other theory of quotation, can account for why 'Waiter, there's "A FLY" in my alphabet soup' is amusing (this is adapted from a cartoon in *Private Eye*, 16 November 1984).

Hong Kong University,
Pokfulam Road,
Hong Kong

© LAURENCE GOLDSTEIN 1985

REFERENCES

[1] D. K. Buckner, 'Goldstein on Quotation', *Analysis*, 44 (1984), 189-90.
[2] D. Davidson, 'On Saying That', *Synthese*, 19 (1968-9), 130-46.
[3] D. Davidson, 'Quotation', *Theory and Decision*, 11 (1979), 27-40, reprinted in his *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press, 1984, to which latter my page references are made.
[4] L. Goldstein, 'The Reasons of a Materialist', *Philosophy*, 55 (1980), 249-52.
[5] L. Goldstein, 'Scientific Scotism – The Emperor's New Trousers or Has Armstrong Made Some Real Strides?', *Australasian Journal of Philosophy*, 61 (1983), 40-57.
[6] L. Goldstein, 'Pure Categorial Principles', *The Monist*, 66 (1983), 410-21.
[7] L. Goldstein, 'Quotation of Types and Other Types of Quotation', *Analysis*, 44 (1984), 1-6.
[8] L. Goldstein, 'The Paradox of the Liar – a Case of Mistaken Identity', *Analysis*, 45 (1985), 9-13.
[9] W. V. Quine, *Mathematical Logic*, Cambridge, Mass.: Harvard University Press, 1940.
[10] J. Searle, *Speech Acts*, Cambridge: Cambridge University Press, 1969.
[11] J. Searle, *Intentionality*, Cambridge: Cambridge University Press, 1983.
[12] A. Tarski, 'The Concept of Truth in Formalised Languages' in his *Logic, Semantics and Metamathematics*, Oxford: Clarendon Press, 1956.

INDETERMINATE PEOPLE

By HUGH S. CHANDLER

ARE there physically possible cases of non-epistemic indeterminacy with respect to personhood and the possession of a mind? I think there are, and that they help us understand the properties of being a person and having a mind as well as providing an interesting case of indeterminacy of translation.

I

My creatures belong to the family invented by John Searle (see 'Minds, Brains, and Programming', *The Behavioral and Brain Sciences* 3, pp. 417–24). One member of this family, you may recall, is a composite entity made up of a complex electronic and mechanical vehicle, plus an homunculus who does not speak Chinese (but does, apparently, speak English). The composite entity 'acts' and 'talks' like a native Chinese speaker. The trick is done by the homunculus working a program in number code. Searle's intuition is that the composite entity does not understand Chinese.

What sort of argument underlies that intuition? References to composite entities can sometimes be unpacked. For example, someone says 'That truck wants to pass us'. Here, no doubt, 'that truck' denotes a composite entity made up of an inanimate vehicle and a human driver. The predicate 'wants to pass us' is applicable to the driver and not to the inanimate vehicle. So, we might think, if some mentalistic property is correctly ascribed to the truck, it must be just because it is a property of the driver. By contraposition, if the driver does not want to pass, it cannot be true that the truck wants to pass. Similarly, in the case of the Chinese-spouting composite entity one might suppose that any understanding, any mental activity, performed by the entity can only be performed by its homunculus. The homunculus does not understand, or speak, Chinese. Hence the composite entity doesn't either.

The argument is obviously hazardous. Imagine animals of two sorts that have evolved in symbiotic union. One is, as it were, the 'brain' of the pair, and the other is a mindless, well-armoured, lobster-like creature. The brainy members only emerge from their lobster bodies once a year (on a dark night) to mate with others of their kind. They have no language of their own. On the other hand, the body-brain pairs (the 'duos') use a rich language in dealing with other such pairs. Here I think we should say that the brainy animal does the duo's thinking, hoping, fearing, and so forth. Nevertheless, it does not follow that the brainy creature must, on its own, understand duosian language.

In any case, neither Searle's composite entity nor the duos are the creatures we are seeking. We want something like the following. Imagine that thirteen species of animals have evolved together as a symbiotic cluster. The composite entities thus formed can 'speak' and 'act' as rational beings. Let's suppose that the components are individually almost as intelligent as the composite creatures. They have languages of their own, and, during certain seasons, live with their own species. The composite entities which they form during the remainder of the year (the 'tredecims') interact in more complex ways – for example they construct apartment houses and compose music. What do the member-units recollect of their stint as components of, say, a house-building tredecim? They say they were engaged in difficult 'computational' tasks. It seems clear that there were at least thirteen minds at work; but were there *fourteen*? Does a tredecim have a mind of its own over and above the minds of its components?

Here is Tredecim One (T-1). Sometimes I see it as a creature with its own mind, and sometimes I see it as a sort of committee or team – its components have minds, but it's just a lump. Is there something crucial hidden here? Something totally inaccessible? We can readily discover what sort of person T-1 is, its virtues and vices, its hopes and fears – insofar as it is a person at all. But how can we discover whether it is a person or a team? Does God know?

Imagine T-1's disintegration into components. The process causes, or seems to cause, terrible pain for T-1; but no component feels pain. So is the process painful or not?

Thomas Nagel writes: '... if things emerged from a spaceship which we could not be sure were machines or conscious beings, what we were wondering would have an answer even if the things were so different from anything we were familiar with that we could never discover it. It would depend on whether there was something it was like to be them, not on whether behavioral similarities warranted our saying so ...' (*Mortal Qestions*, pp. 192–3). No doubt he would make the same claim in regard to T-1. But is there something it is like to be a tredecim, or is there not? On Nagel's theory we are confronted by a transcendental, totally inaccessible, fact. (Nagel sides with Leibniz against Locke on this issue. See the abridged edition of Leibniz's *New Essays*, edited by Peter Remnant and Jonathan Bennett, pp. xxxiv–xxxvi.) It seems to me preferable to hold that the matter is non-epistemically indeterminate. Unfortunately I have no proof that my preference is the right one.

II

Would borderline 'people' like the tredecim show that the property of being a person is somehow unreal – not the sort of property God recognizes? It might be useful here to separate two sorts of classes that can have borderline 'members'.

Potentially less fuzzy classes

In some cases a fuzzy term can, and should, be made more precise in some particular way. Thus, for example, the denotation of the term 'ape' has been considerably sharpened over the last two hundred years. The motive for this sharpening is the quest for natural kinds. There is, of course, residual vagueness. Perhaps any further sharpening would be arbitrary.

I suppose there are other fuzzy terms that could be made more precise, but which are such that arbitrariness would set in immediately. Thus we could fix a fairly narrow height range above which a building is a 'sky-scraper'. There are, in fact, no end of height ranges we could equally well choose for this purpose.

Some realists like Hartry Field take terms for potentially less fuzzy classes to partially designate each of the minimally fuzzy classes they could designate if properly sharpened (see 'Quine and the Correspondence Theory', *The Philosophical Review* 83, April 1974, pp. 200–28). The idea is certainly plausible. As an abstract example, think of a class term 'N', which seems to designate a class having two clear members, *a* and *b*, and two borderline 'members' *c* and *d*. 'N' may best be understood as partially designating each of four precise classes: (1) *a* and *b*, (2) *a*, *b*, and *c*, (3) *a*, *b*, and *d*, and (4) *a*, *b*, *c*, and *d*. A term that can be sharpened in such a way as to designate a real kind, or that can be sharpened in various equally arbitrary ways, does not, prior to such sharpening, determinately designate any class. It *partially* designates various natural, or factitious, classes.

Fuzzy physical classes

Consider the class whose members are just two particular electrons, *a* and *b*. Now *a* interacts with a third electron, *c*, and two electrons, *d* and *e*, emerge from the interaction. Electrons obey Fermi statistics, and this entails, under the usual interpretation, that it is demonstrably wrong to identify *a* with either *d* or *e*, and also wrong to *distinguish* *a* from *d* or from *e*. Thus '*d* is *a*', and '*e* is *a*', are demonstrably neither true nor false. It is intrinsically indeterminate whether or not *d* and *e* are members of the class whose members are just *a* and *b*.

The case is confusing. But it suggests what it would be like to find physical (primary) fuzziness. What seems to be required is proof that any reduction of the fuzziness would lead us to make false predictions. The case also indicates that the occurrence of a borderline 'member' of a class cannot, by itself, prove the class unreal.

What about the property of being a person? I take the tredecims to show that the property is much fuzzier than it needs to be. Our classificatory problem in regard to the tredecims is linked to our worry that their components are not sufficiently committed – biologically, or in terms of time and effort – to tredecim life. They have too much of a life of their own. There are various possible cases here. We can imagine components that are more committed, and components even less committed. (Imagine, for example, a team of ordinary humans who together work a 'person' program on Saturday afternoons as a game.) The boundary between people and non-people is very vague in this region. But nothing prevents us from sharpening the denotation of the term 'person'. We could come to hold that the tredecims are people, or that they are not. Such sharpenings might be impractical and morally repulsive, because of the trivial differences we would then take as decisive; but we would not be forced into demonstrable absurdities. Consequently, at least on the realist view sketched above, we should hold that the term 'person' partially designates various classes and does not *determinately* designate any class.

III

Can we understand the language of the tredecims ('Tredecimese')? Can we translate it into English? To make these questions more specific, suppose T-1 'says':

(U1) 'Iyay amyay inyay ainpay.'

Might that mean 'I am in pain'?

Theory I

Whether or not tredecims have minds is irrelevant to the *meaning* of their utterances. Tredecim utterances entailing that some tredecim (e.g. the 'speaker') has a mind are either indeterminate or false in *truth value*.

The argument for this approach would be that meaning is fixed by such things as observable uses of utterances and logical relations among sentences. This being so, nothing prevents Tredecimese from being translatable into English (*via* the usual procedure for decoding pig-latin). Suppose, then, we take (U1) to mean 'I am in pain', because of its observable uses and so forth. Are we to think that (U1) may well be *true*? If 'I am in pain' is true, then the speaker has a mind. (The first person singular implies that the speaker is not a team. Similarly, I imagine, (U1) implies that the 'speaker' is not an eamtay – whatever *that* is.) Given that whether T-1 has a mind or not is non-epistemically indeterminate, we must hold that (U1) is, at best, indeterminate in truth value.

This theory is deficient in charity. After all, (U1) may be a sentence in frequent use among tredecims. Are we to insist that in its normal employment it can never be true?

Theory II

Tredecim utterances that might be interpreted as entailing that some tredecim has a mind are normally *true,* but are *indeterminate* in *meaning*, relative to our language. Thus, for example, it is equally incorrect to translate (U1) as

(1) 'I am in pain' or
(2) 'We are hurting'

(where 'hurting' is a state in which either one or more members of a group are in pain, or the group itself, but no member, is suffering damage).

Presumably (U1) is exactly like 'I am in pain' in regard to its observable uses and its logical relations to other sentences; but we need not conclude that the two sentences have the same meaning. According to theory II, (U1) has no correct translation into English.

Section II included the suggestion that we could decide the tredecims are people. If we decide to count them as people, we should translate (U1) as 'I am in pain' (and assume it to be true). If we decide they are not people, we must take (U1) to mean something like 'We are hurting'. Our freedom to sharpen the concept 'person' in either of these ways, or leave it as fuzzy as it now is, compels us to be non-realists in regard to the meaning of many tredecim utterances.

University of Illinois at
Urbana-Champaign,
810 South Wright Street,
Urbana, Illinois 61801, U.S.A.

© HUGH S. CHANDLER 1985

ARE MENTAL EVENTS IN SPACE-TIME?

By P. F. GIBBINS

SUPPOSE, as no one does, that one's feeling a twinge is an event in space-time which is (strangely) not part of the world-line of one's brain but which is somewhere else. Concede to our ordinary scientific prejudice that there are no twinges without brain-processes.

Then in frames of reference moving with respect to one's brain the twinge will not be simultaneous with the twinge-making brain process.

Even the dualist wants to keep mind and brain in step. And nowadays he will want to keep on the right side of the theory of relativity. So if twinges are in space-time they must be coincident with their twinge-making brain-processes. This result is the contribution of the lorentz transformation to the philosophy of mind. For on the galilean view, twinges and twinge-making brain processes that are simultaneous in one frame of reference are simultaneous in all frames of reference. Call this *the argument from simultaneity*. If mental events are simultaneous with brain processes then if they are in space-time, they are coincident with their corresponding brain-processes. Relativity is an obstacle for the galilean dualist who imagines that mental events inhabit a non-spatial realm just like classical space-time but with the space projected out.

Michael Lockwood (ANALYSIS 44.1, January 1984, pp. 22–5) supplies what one might think of as the dual of the argument from simultaneity: since events which are temporally separated in one frame of reference are spatially separated in others it follows that mental events are spatial as well as temporal 'since events could hardly be spatially separated without being spatially located' (Lockwood, op. cit., p. 23). He also supplies a method by means of which one can locate mental events in space-time, given that that is where they are: assuming causality, if a physical event A causes a mental event B which in turn causes a physical event C then B must lie in the intersection of the forward light-cone from A and the backward light-cone from C.

Note that all these arguments and the space-time diagram (Lockwood, op. cit., p. 24) that accompanies them depend on the assumption that mental events are in space-time. None shows that mental events are in space-time. Lockwood can infer spatial separation in one frame of reference from temporal separation in another because he assumes that mental events have a location in any given frame.

Arguments very similar to Lockwood's occur in Robert Weingard's 'Relativity and the Spatiality of Mental Events', *Philosophical Studies* 31 (1977), pp. 279–84. Like Lockwood, Weingard uses a space-time diagram (oddly drawn) to infer that if special relativity applies to a datable event then the event is in space as well as time. But the diagram presupposes that mental events are in space-time. Weingard's argument for the antecedent, that special relativity applies to mental events (odd phrase!), is (again) that they share the field of temporal relations with physical events.

The obvious circularity in the Lockwood–Weingard thesis conceals what must be their real motivation. That is

> [t]he fact that mental events admittedly have temporal relations has much force, now that time and space are so much less distinct than they were.

> It has become difficult to hold that mental events, though in time, are not in space.

Or so Bertrand Russell put it, with uncharacteristic caution in 1926 (see *The Analysis of Matter*, Allen and Unwin 1927, p. 384). In other words, relativity – special and general – so mix up space and time that dualism becomes incoherent. (See also Lockwood's reply to Gordon in ANALYSIS 44.3, June 1984, p. 128.)

But can 'mental events belong to the same temporal order as physical events' (Lockwood ANALYSIS 44.1, January 1984, p. 23) and yet not be in space? Is there a way out for the stubborn dualist?

I think there is, at least for the *lorentzian* dualist who agrees with Russell that relativistic metaphysics has much force but who denies that it has enough force to sort out the philosophy of mind.

The lorentzian dualist, whom I want to defend against the charge of incoherence, can be just like the galilean dualist but with a modified view of the time in which mental events occur, this forced by the fact that mental events have causal relations with physical events with which they keep in step. For him the relativity theories are theories about space-time and so apply directly to only *some* of the entities which can stand in temporal relations, namely space-time events.

The temporal results of relativity will apply to mental events *only indirectly* via the space-time events with which they are associated. Like the galilean dualist the lorentzian dualist holds that the relation of temporal order – say 'after' – on causally connectible mental events is asymmetric and transitive. Unlike the galilean dualist he holds that the relation of simultaneity on causally connectible mental events, though symmetric, is not transitive. But there can be pairs of mental events which are neither simultaneous nor temporally ordered, so that mental events associated with (caused by or causing) space-time events that are not causally connectible – such as brain-processes separated by a space-like interval – are themselves temporally incomparable.

For the updated, lorentzian, dualist, who is no dottier than his galilean predecessor, there are therefore no superluminal telepathic communications which can have physical effects either directly or indirectly. And perhaps there are no such things at all.

Department of Computer Science,
University of Bristol,
University Walk, Bristol BS8 1TW

© P. F. GIBBINS 1985

EINSTEIN, GIBBINS AND THE UNITY OF TIME

By MICHAEL LOCKWOOD

GIBBINS (this issue, pp. 145-7) thinks that there is a circularity in the arguments Weingard (*Philosophical Studies* 31, 1977, pp. 279-84) and I (ANALYSIS 44.1, January 1984, pp. 22-5, and 44.3, June 1984, pp. 127-8) present for the conclusion that mental events are in space. His contention is that we, in effect, presuppose from the outset that mental events are in space-time – a presupposition that is reflected in our use of space-time diagrams.

As regards the diagrams, at least, the charge of circularity may be immediately rebutted. My own diagram does indeed presuppose that mental events are in space-time. But then it is not intended (except perhaps obliquely) to have any role in helping to *establish* that conclusion; I take the conclusion already to have been established in the first part of the article. Rather, it illustrates the second of my two claims: namely that, assuming mental events to be in space-time, special relativity (in conjunction with causality and psycho-physical causal interaction) yields a method of actually determining the spatial location of a mental event. (To be sure, I intend this to reinforce the earlier conclusion, but only in the sense of allaying any verificationist misgivings someone might have about a view which held mental events to be in space, without being able to provide any method of actually locating any given mental event.)

By contrast, Weingard's article does not presuppose that mental events are in space-time, since it makes no reference to mental events; rather it is being used to help establish a quite general lemma (his L1): that '[i]f special relativity applies to an event and that event occurs in time, then it occurs in space (i.e. it is located in a definite spatial place)' (op. cit. p. 279). That special relativity does apply to mental events has, on his view, to be independently argued for (as his lemma L2).

What, then, of the premises to which Weingard and I appeal, by way of arguing that mental events are in space? They are, respectively, (P1) that special relativity is correct and (P2) that mental events are in time – I gloss this in my own article by saying that mental events belong to the *same* temporal order as physical ones (since one cannot exclude *a priori* the existence of a plurality of mutually disjoint sets of events within each of which temporal relations obtain, but between which they do not). The argument from P1 and P2 to the spatiality of mental events is then, in a nutshell, that if special relativity is correct then any event that is in time – or more strictly, any event that belongs to the same temporal order as physical events – is in space-time, and any event that is in space-time is also in space. I cannot for the life of me see how this argument can be supposed to 'depend on the assumption that mental events are in space-time' (Gibbins, p. 146). On the contrary, Weingard and I take this to follow, logically, from premises P1 and P2, the assumptions that they are in time (the same temporal order as physical events) and that time and space are related as special relativity says that they are. Perhaps Gibbins thinks that the argument will not go through unless tacit appeal is made to the additional premise that mental events are in space-time; but if so, I am at a loss to see why.

Indeed, Gibbins' charge of circularity sits ill with what he then goes on to say. For the way out that he offers the dualist seems to involve the possibility of denying that premise P2 is in fact *true* in the sense required to establish the conclusion. If I understand Gibbins' suggestion, what is being envisaged is the possibility of distinguishing two senses in which events might be held to be in time – a *primary* sense that applies directly to physical events, and a *derivative* sense that applies to any events that stand in a causal relation to events that are in time in the primary sense. The dualist might then insist that the correctness of the special theory of relativity implies only that events that are in time in the primary sense are also in space, whereas it has not been established that mental events are in time in anything but the derivative sense.

Interestingly enough, Weingard explicitly considers a generalized form of this line of argument in his own article (op. cit. pp. 283–4); instead of 'causal relation' he says 'non-temporal association'. What he says in rebuttal is:

> [I]t is simply a fact of our experience that we are aware of mental events directly bearing temporal relations to (non-mental) physical events, and not of (non-mental) physical events non-temporally associated with the mental event being temporally related to other (non-mental) physical events. When I am aware that I have a pain and that it is occurring before lunch, it is, I think, an obvious fact of experience that it is the pain itself that I am aware of occurring before lunch, and not that some (non-mental) physical event the pain is non-temporally associated with is occurring before lunch (ibid. p. 284).

There is, I fear, something wrong with this argument. Suppose someone says, for example, 'When the train had pulled out, I suddenly felt very much alone'. Then what the speaker no doubt takes himself to have been aware of is that his feeling directly stood to the train's departure in the temporal relation 'very soon after'. But, for all that Weingard says, it is presumably open to Gibbins' dualist to deny that what he was *in fact* aware of was the holding between these events of the very same primary relation of temporal ordering as is to be found amongst physical events, and which enters into the mathematical statement of physical law. Compare someone who claimed that he was aware of a bright yellow after-image directly bearing to the lamppost the spatial relation 'to the left of'. Clearly, this awareness (or pseudo-awareness) does not license the attribution to after-images of spatial relations with physical objects in the same (primary) sense in which physical objects may be said to stand in such relations to one another.

I doubt whether there is, in fact, any simple knockdown way of refuting Gibbins' proposal. But I cannot agree with him that his 'lorezian dualist ... is no dottier than his galilean predecessor'. The position of the galilean dualist was in conflict neither with science nor with common sense. The lorenzian dualist, by contrast, surely is in conflict with common sense. For it is part of our common-sense view of reality that mental events stand in temporal relations in precisely the same primary sense as do physical events. Indeed, there is surely something right about Kant's observation that time is the form of the inner sense: what we are aware of within our own stream of consciousness serves, in a sense, as a paradigm of temporal relatedness, which is then extended to the world at large. Of course, the lorenzian dualist might concede that his view conflicts with common sense in this regard but, accepting the *validity* (as opposed to soundness) of the argument Weingard and I have presented, defend this as the only way of avoiding the offence to common sense involved in attributing spatial locations to mental events. But I should have thought that locating mental events in space was far less offensive to common sense than denying them, in the primary sense, a location in time.

Just how perverse the lorenzian dualist's position is becomes even more evident when one reflects that any set of events could, using Gibbins' strategy, be exempted from the strictures of the special theory of relativity: all one has to say is that such ostensible temporal relations as they stand in hold merely in a derivative sense, grounded in the causal relations in which they stand to such events as are not so exempted. It is not, I think, merely the arbitrariness of such a move that offends. Rather, there is something deeply perplexing about so divorcing temporal relations, in their central, primary sense, from causal relations. Time and causality are, in our ordinary ways of thinking, intertwined and interdependent concepts.

Finally, I am grateful to Gibbins for drawing my attention to Weingard's article; I was previously unaware of its existence.

Department for External Studies,
Oxford University,
Wellington Square, Oxford OX1 2JA

© MICHAEL LOCKWOOD 1985

A NOTE ON TEMPORAL PARTS

By H. W. NOONAN

I HOLD that *if* enduring objects have temporal parts then it is reasonable to say that (some of) the temporal parts of an enduring object of sort S are themselves objects of sort S. If persons have temporal parts it is reasonable to say that some of these are themselves persons, if tables have temporal parts it is reasonable to say that some of these are themselves tables, and so on.

But I find that this view encounters resistance, not only (and perhaps somewhat inconsistently) from the opponents of the temporal part metaphysic, but also from its proponents. In what follows, I defend it.

Sometimes, I find, it is suggested that maintaining that temporal parts of persons are themselves persons is on a par with maintaining that their spatial parts are themselves persons. But this is not so. Of course my left foot is not a person: it is the wrong shape, size and structure and has none of the relevant psychological properties (in fact, it has *no* psychological properties). But the same overwhelming case cannot be brought against the suggestion that Noonan-in-1984 is a person. For it is common ground among proponents of the temporal part metaphysic that (as David Lewis puts it, *Philosophical Papers*, Vol. I, Oxford University Press, 1983, p. 76):

> A person-stage is a physical object just as a person is. (If persons had a ghostly part as well, so would person-stages.) It does many of the same things that a person does: it talks and walks and thinks, it has beliefs and desires, it has a size and shape and location. It even has a temporal duration ... It begins to exist abruptly, and it abruptly ceases to exist soon after. Hence a stage cannot do everything that a person can do, for it cannot do those things a person can do over a longish interval.

(Not all person-stages will satisfy Lewis's description. Some will be too short-lived. It is only of those that do that I claim that they can be reasonably described as persons.)

It might be said that person-stages cannot be persons just *because* they are parts of persons. But the principle this argument appeals to, namely that the parts of something falling under a certain sortal concept cannot themselves fall under that sortal concept, is dubious when applied to spatial parts (think of Wiggins' example of the Pope's three crowns which are one crown), so it can hardly merit the status of an unargued assumption in the case of temporal parts.

It might be said that nevertheless person-stages cannot be persons, since they do not last long enough. But we do not make it a requirement on something's being a person that it have a certain temporal size (duration) and of course many people die at birth or soon after.

Again, it might be said that person-stages cannot be persons because of the manner in which they begin to exist and cease to exist. But let us imagine, varying a suggestion of Wittgenstein's, that something happens which we are tempted to describe as a person's coming suddenly into existence, going about his business, and then ceasing to exist as abruptly as he began. Would it really be straightforwardly *wrong* to give such a description, or would the sensible thing not just be to say, with Wittgenstein, 'say what you like so long as you are clear about the facts'? (Certainly Lewis' could not say that this description was straightforwardly wrong, see his *Philosophical Papers*, Vol. I, p. 76: 'it is possible that a person stage might exist. Suppose it were to appear out of thin air, and vanish again. Never mind if it is a stage *of* any person (though in fact I think it is).')

Now the facts about person-stages (I am assuming, since I am taking a temporal part metaphysic for granted) are those outlined in the passage from Lewis. And my suggestion is just that given *these* facts it is not straightforwardly wrong to describe person-stages as persons – we have no nice strict set of necessary and sufficient conditions for personhood that disqualifies them (indeed, on one well known definition, John Locke's, it is strictly correct that they are persons). Whether it is reasonable for the temporal part metaphysician so to describe them must therefore turn on other considerations. In particular, I believe, it must turn on whether by doing so he puts himself in a better position to answer his opponents. Now elsewhere I have argued that this is so (*Objects and Identity*, Martinus Nijhoff, 1980), and so I do accept that it is reasonable to describe temporal parts of persons as themselves persons (and temporal parts of tables as themselves tables, and so on).

Of course, if this view is not to be easily refutable it must come as part of a package. In particular, in order to make sense of such statements as Lewis's '. . . a stage cannot do everything a person can do' we need a distinction between restricted and unrestricted quantification and we must maintain that 'some S is so and so' (for sortal S) is not simply equivalent to 'something is an S and is so and so'. But not only is this view a respectable one, it is also an essential ingredient in Lewis's counterpart theory translation scheme (see his 'Counterparts of Persons and their Bodies', *Philosophical Papers*, Vol. I) and, I believe, an essential ingredient in *any* defensible temporal part metaphysic (for the argument for this see my 'The Closest Continuer Theory of Identity', forthcoming in *Inquiry*).

Another view we must accept if we are not to be open to easy refutation in maintaining that the temporal parts of persons are themselves persons is that counting is not always in accordance with identity *simpliciter* (otherwise we would have to accept that there are many more people than we think that there are). But this view too should be congenial to a temporal part metaphysician, and again is explicitly accepted by Lewis.

The University of Birmingham,
P.O. Box 363, Birmingham B15 2TT

© H. W. Noonan 1985

SWAIN ON THE BASING RELATION

By JONATHAN L. KVANVIG

SUPPOSE we want to know whether a person justifiably believes a certain claim. Further, suppose that our interest in this question is because we take such justification to be necessary for knowledge. To justifiably believe a claim requires more than there being a justification for that claim. Presumably, there is a justification for accepting all sorts of scientific theories of which I have no awareness; because of my lack of awareness, I do not justifiably believe those theories. Further, even if I, by chance, did believe one of those theories, I might not justifiably believe it; for I may be completely unaware of any justification for it. I then would not believe it *because of* the justification for it.

One plausible account of what it is to believe because of the evidence is the causal account. On the causal account, what distinguishes one's *having* reasons from there *being* reasons is for the reasons to be causally responsible in some way for the belief in question. Such a theory, it is claimed, adequately explains how one can have a belief which is irrational even though there are good reasons for the content of that belief; the belief might be irrationally held because the evidence might not be causally responsible for the belief.

The context of this discussion is the context of determining what sort of conditions a belief must meet in order to meet the justification condition for knowledge, and in this context, two sorts of causal theories might be proposed. One sort requires that the evidence prompt the original acquisition of the belief in question. The other sort requires that the evidence sustain the holding of the belief at the particular time in question. Marshall Swain has offered a theory which includes elements from both sorts of causal theories, for he claims that both the acquisitional causes and the present sustaining causes are relevant to the justificatory status of the belief (Marshall Swain, *Reasons and Knowledge*, Cornell 1980, pp. 98–9). Further, according to Swain, causal basing is sufficient, though not necessary, for a belief to be accepted because of, or to be based on, the justification for it (ibid. p. 86). The reason for not requiring actual causal relations concerns a counterexample which has been proposed that supposedly undermines that sort of theory. The counterexample concerns a gypsy lawyer who, like the rest of his contemporaries, takes his client to be guilty. However, because of his gypsy nature, the lawyer is inclined to trust what the tarot cards say, and upon learning that the tarot cards say that his client is innocent, comes to believe that his client is innocent. What the tarot cards say also prompts the lawyer to re-examine the evidence, which the lawyer comes to recognize conclusively establishes that his client is innocent. However, given his rather impressionable character, the lawyer also realizes that were the sustaining power of the tarot cards removed, the sway of public opinion would cause him to be unable to see that the evidence establishes his client's innocence. Nonetheless, the lawyer now justifiably believes that his client is innocent on the basis of his examination of the evidence. But this examination of the evidence neither prompts his belief that his client is innocent, nor does it sustain his belief that his client is innocent – his belief in what the tarot cards say holds the dubious distinction of being responsible for both (Keith Lehrer, *Knowledge*, Oxford 1974, pp. 124-5).

Causal theorists have had two sorts of reactions to the gypsy lawyer case. Some have denied that the gypsy lawyer case is really a counterexample by denying that the lawyer justifiably believes that his client is innocent (cf., for example, Robert Audi's recent paper, 'The Causal Structure of Indirect Justification', *Journal of Philosophy* 1983, pp. 398-415). The other response is Swain's – to grant that the lawyer justifiably believes that his client is innocent and attempt to save the heart of the causal theory in spite of this admission. Swain insists that the lawyer only justifiably believes the claim in question because (roughly) his belief that the evidence establishes the innocence of his client would have caused the belief were the actual cause absent. Thus, Swain claims, the heart of the causal theory can be maintained by insisting that if the examination of the evidence does not *actually* cause belief, it must *counterfactually* cause belief. One might capture this suggestion as follows:

> S1: Person A bases his belief that p on reasons R if and only if for any r which is a member of R, (i) r is a belief state or A is directly aware of r; and (ii) *A's having r* is either an actual or a counterfactual cause of *A's believing that p*.

The phrase 'counterfactual cause' is ambiguous here, though, for there are many ways in which one event can be a counterfactual cause of another. For example, suppose that e actually causes p, but that f would have caused p if e had not – f then counterfactually causes p. But further suppose that if neither e nor f had caused p, g would have – g too is a counterfactual cause of p. And there seem to be as many counterfactual causes here as one would wish.

Let us order these counterfactual causes by the relation 'buried by'. One counterfactual cause is buried by another if the second must fail to obtain in order for the first counterfactual cause to be the actual cause. So, in the example of the above paragraph, g is buried by f because f only requires that e not be a cause of p in order for f to be a cause, whereas g requires that both f and e not be a cause of p in order for g to be a cause.

Swain's view is that the only sort of counterfactual causes appropriate are those that are not buried by any other counterfactual causes. Thus, we can capture his view as follows:

> S2: Person *A* bases his belief that p on *R* iff for every *r* in *R*, (i) *r* is either a belief state or *A* is directly aware of *r*, and (ii) *A's having r* is either a cause of *A's believing that p*, or *A's having r* is a counterfactual cause, buried only by an actual cause, of *A's believing that p* where both *A's believing that p* and *A's having r* occur (Swain, op. cit. pp. 74, 86-7).[1]

What prompts this move to counterfactual causes, to repeat, is the gypsy lawyer case in which the lawyer is aware of the evidence and how it confirms his belief, but this awareness of the evidence and what it shows is not causally responsible for the belief. But why should one stop at counterfactual causes buried only by the actual causes? For example, suppose that we altered the gypsy lawyer case like this: not only does the gypsy lawyer trust the tarot cards, he is also given to visiting fortune-tellers. Further, he has not only consulted the tarot cards about his client, he has also visited his favourite fortune-teller who has informed him that he will win his case because his client is really innocent. Now, the lawyer's belief in what the fortune-teller says is not causally responsible for his belief in his client's innocence; the tarot cards hold that honour. But the lawyer's belief in what the fortune-teller says would have caused the belief if the tarot cards had not, i.e. the belief in what the fortune-teller says is a counterfactual cause, buried only by the actual cause, of belief that his client is innocent. What the fortune-teller says would not only have caused him to believe that his client is innocent, but would also have prompted him to re-examine the evidence. But, just as in the original case, the sway of public opinion would have resulted in his failure to appreciate the force of the evidence, were the influence of the fortune teller removed.

The same reasons for thinking that the lawyer justifiably believes that his client is innocent in the first case are reasons for thinking that he justifiably believes that his client is innocent in this expanded case. But, contrary to S2, his examination of the evidence is not a counterfactual cause of belief buried only by the actual cause of belief. For, in this case, the examination of the evidence is at best buried by both the actual cause and one counterfactual cause.

Solving this problem will not be simple, for it is easy to see how to extend the above counterexample to allow third level counterfactual causes (i.e. causes buried by two counterfactual causes) to satisfy the requirements of the basing relation, and then on to fourth level, etc. So, suppose Swain tries:

> S3: *A* bases his belief that p on *R* iff for every *r* in *R*, (i) *r* is either a belief state of *A*, or *A* is directly aware of *r*, and (ii) for the set *C* of possible counterfactual causes of belief (a) either

---

[1] My notion of a counterfactual cause buried only by the actual cause is just Swain's notion of pseudo-overdetermination. I should point out, though, that I ignore some complications in Swain's account that are not relevant to my discussion.

A's having r causes the belief that p, or (b) *A's having r* is a member of *C*.

There is an ambiguity in S3 concerning what is properly allowed as a member of *C*. We could allow either event-types, which may or may not have occurrences, or we can restrict set membership to occurrent events. In Swain's original theory, he understands events as occurrent events (ibid. p. 46), so suppose we consider that option first.

There is no reason to think that the gypsy lawyer's appreciation of the force of the evidence would cause his belief even if no other occurrent event were to have caused it and yet both this appreciation and the belief in question were to occur. For it is surely possible that the lawyer's appreciation of the force of the evidence would be a cause of the belief in question only if no other occurrent event were to be a cause *and* some other conditions were to fail to obtain, where these conditions do not presently obtain. For example, let us suppose that the only occurrent event which buries the causal activity of the lawyer's appreciation of the evidence is the lawyer's visiting the fortune-teller. The question is whether the lawyer's appreciation of the evidence is a second level counterfactual cause or not, and the answer is that it need not be. For it is possible that, were the lawyer not to have visited the fortune-teller nor consulted the tarot cards, yet believed his client was innocent and appreciated the force of the evidence regarding his client's innocence, his appreciation of the force of the evidence would not have caused his belief. How could this happen? Perhaps he would have read the *National Enquirer*, which carried a story which would have both caused him to believe that his client was innocent and caused him to re-evaluate the evidence. Hence, even if his appreciation of the evidence is only buried by one counterfactual cause (where only occurrent events are allowed in the class of counterfactual causes), it does not follow that were the burying cause (and what buries it) absent, the buried cause would have been the actual cause.

Further, even if the lawyer's appreciation of the force of the evidence were not a counterfactual cause, where only occurrent events are allowed in the class of counterfactual causes, there is no reason to think that the lawyer does not justifiably believe that his client is innocent on the basis of his examination and appreciation of the force of the evidence – at least no reason that would not also imply that in the original lawyer example, the lawyer justifiably believes that his client is innocent.

The lesson here is that if we are to preserve S3, we cannot limit the class of counterfactual causes to occurrent events. Yet, if we allow non-occurrent events in that class, an equally vicious problem arises, for if it is (logically) possible that the lawyer's appreciation of the evidence cause the belief, then clause (ii)(b) of S3 begins to look very much like a necessary truth. That clause claims that were both the belief and the appreciation of the evidence to occur and were nothing else to have caused the belief, then the appreciation of the evidence would have caused the belief. Since we are allowing non-occurrent events to be included in the 'nothing else' clause, (ii)(b) could escape being a necessary truth only if it is possible that the belief in question be uncaused, for if the belief cannot be uncaused and if nothing but the appreciation of the evidence is a cause, then the appreciation of the evidence *must be* a cause.

It may be possible for beliefs to exist uncaused, but if it is, we at least need an argument that it is so; but even if such an argument is forthcoming, that is not sufficient for defending S3. For not only must uncaused beliefs be possible to preserve S3; Swain's view requires that if such uncaused beliefs are possible, none of them can exist uncaused and yet be justifiably held on the basis of evidence of which the person is aware. Perhaps it is possible that God create a world in which there is only one causal relation: that relation between his activity and the existence of the world. Further, it may be possible that some person exists in such a world and has beliefs which are uncaused. But, if all of this is possible, I cannot see why it is not also possible that that person justifiably believes that he exists on the basis of his awareness that he thinks; yet this is just what a defender of S3 must hold. He must hold that uncaused beliefs are possible, and also that none of them can be justifiably held on the basis of other beliefs or awarenesses. To understate the point, it is a little difficult to see what sort of argument there could be for such a view.

Thus Swain's causal theory of the basing relation faces severe difficulties. First, if some counterfactual causes are allowed, there is no good reason to restrict the class to those buried only by the actual cause. Further, not even allowing any counterfactual cause, where the class of causes in question is composed of occurrent events, is sufficient to rescue his view; for the gypsy lawyer's appreciation of the evidence might not be a counterfactual cause at all on this construal, and yet there is as much reason to think he justifiably believes that his client is innocent as there is in the original gypsy lawyer case. Finally, the theory cannot be rescued by allowing non-occurrent events in the class of counterfactual causes, for then to claim that the appreciation of the evidence would have caused the belief had both of these occurred and nothing else caused the belief, sounds very much like a necessary truth. The only way to keep it from being a necessary truth is to deny that all beliefs have causes; but to do this, it would seem, gives up the heart of the causal theory. For if beliefs can exist uncaused, there does not seem to be any reason for thinking that they cannot exist uncaused and be justifiably held. Of course, I do not wish to rest the plausibility of rejecting Swain's causal theory on the possibility of uncaused and justifiably held beliefs; my point is simply that I can see no reason for not thinking that if uncaused beliefs are possible, so are uncaused and justifiably held beliefs.

My purpose has been to show that if one, like Swain, accepts that the gypsy lawyer knows that his client is innocent, no sort of causal theory of the basing relation can be adequate. One cannot preserve either the acquisitory or sustaining condition in the face of the gypsy lawyer case; and one cannot preserve either an actual or counterfactual causal view in the face of other cases as reasonable to accept as the original gypsy lawyer case. The only option, then, for the causal theorist is to deny that the gypsy lawyer knows that his client is innocent. I think that claim is clearly mistaken; but that is a topic for another paper. I conclude then that no sort of causal theory, even the exceptionally weak one of Marshall Swain's, can stand in the face of accepting gypsy-lawyer-like cases.

Texas A&M University, ©JONATHAN L. KVANVIG 1985
College Station, Texas 77843-4237, U. S.A.

SHORT ON TELEOLOGY

By A. OLDING

IN a recent paper ('Teleology in Nature', *American Philosophical Quarterly*, Vol. 20, October 1983; page references in the text are to this article) T. L. Short has argued that teleological explanations, of the Aristotelian variety, are still to be found in several areas of natural science, citing in this respect statistical mechanics, the Darwinian theory of evolution by natural selection and I. Prigogine's work on 'dissipative structures'. Leaving aside the work of Prigogine, the surprising thing about Short's argument is that it runs counter to the usual claim that statistical mechanics and the theory of natural selection are paradigms of mechanistic explanation.

First, we had best see what Short thinks a teleological argument is. 'Teleological explanation', he says, 'is explanation by final causes. Final causes are abstract types that so influence processes of mechanical causation that they tend to bring about their own actualization.' (p. 311) He goes on to explain that final causes are introduced 'to explain what mechanical causation by itself cannot explain, namely, the emergence of order from disorder or of uniformity from variety' (ibid.).

So, then, against the conventional wisdom Short maintains that the explanation, in terms of statistical mechanics, of the tendency of isolated systems to reach a uniform temperature ('uniformity from variety'), and the explanation, in terms of natural selection, of the tendency of populations of organisms to evolve, both require fundamental reference to causally potent 'abstract types'. We might wonder just *how* these final causes, taken as abstract types, could be causally potent: surely, we might say, only concrete existents (atoms or galaxies or wombats or whatever) can enter into causal relations. Short thinks not as he shows in a passage where he also tells us just what *are* the final causes in the two cases under consideration.

> The examples of statistical mechanics and the theory of natural selection make it clear how final causes produce finious processes ['finious' is a term invented by C. S. Peirce to refer to end-directed processes]. The type of result toward which the process tends is built into a principle of selection which variations that are retained satisfy ... In the one case it is simply a matter of extremely high probability that any combinations of molecular motions that actually do occur will satisfy the type, 'no decrease in entropy'. In the other case, it is a matter of likelihood that variants that do not satisfy the type 'relatively better suited for reproductive success' will be prevented from reproducing by one or another mechanical event or condition. (p. 314)

With the help of what Short later says in answer to possible objections we can glean two things from this passage. First, the abstract types that are the final causes just are certain principles of selection. This seems to be what is meant when he says that 'The type of result ... is built into a principle of selection, etc.'; and this interpretation is confirmed by what he later says about evolutionary explanations where he claims, 'A principle of selection is the cause given for why variations in inheritable characteristics tend to result in species adapted for survival' (ibid.).

The second point is that what these final causes, as principles of selection, bring about are certain *general tendencies*. What is explained by statistical mechanics is the general tendency towards no decrease in entropy and what the theory of natural selection explains is the tendency towards the survival of the better adapted. The explanations are held to be non-mechanistic just because they are of general tendencies which are accounted for without reference to the particular mechanistic events or processes to be found within a particular exemplification of the relevant tendency. With reference to the theory of natural selection (and the point is similar for statistical mechanics) Short puts the matter as follows:

> But since we predict and explain the general tendency without knowledge of its particulars, and since we know that the tendency would very likely be the same even if the particulars were different (even if other species had evolved instead, they would probably be pretty well adapted for survival too), it follows that the mechanistic explanation of these particular developments is *not* the explanation of the general tendency (ibid.).

It is just here that we must keep our wits about us. I have already suggested that we might raise questions about the possibility of abstract types, whether these are taken as 'principles' or whatever, being causally efficacious. Short's response is to make fun of such questions. We sceptics are narrow-mindedly mechanistic, that is, we think 'that an abstract type could not be a cause unless it got out there and pushed something around – unless, that is, it occupied a point in space and time and exerted a force or was a force exerted from that point' (ibid.). Quite so. I, at least, do believe something like this. But, says Short, this merely begs the question because 'if there is a final causation at all, it cannot be of the same nature as mechanical causation' (ibid.).

Short has a quick way with conceptual thickets but let us ignore that and see how he thinks the issue can be put in a non-question-begging way: '... is any phenomenon explained by reference to something that is not a mechanical cause?' (ibid.). The answer, he thinks, is 'Yes' and it is here that he makes reference to the examples where 'we predict and explain the general tendency without knowledge of its particulars' and where 'we know that the tendency would be very likely be the same even if the particulars were different' (ibid.).

Now in the quotation I gave three paragraphs back Short does not demonstrate that we must believe in final causes – all he shows is that 'the mechanistic explanation of these particular developments is *not* the explanation of the general tendency'. It does not follow that because particular mechanistic causes are not mentioned in explaining the general tendencies of closed systems and evolution therefore some other type of cause must be operating. A rival possibility – I would say *the* rival possibility which only has to be mentioned to become the probability – is that the *details* of mechanistic causes do not have to be mentioned just because the details are not required in explaining the generalized end-states towards which some systems tend. For example, given an isolated inhomogenous mass of gas molecules, all we need to know, *in a general way*, is that they interact according to mechanical principles and then given that, plus statistical reasoning, we can deduce that the gas will become homogenous.

I shall now demonstrate that final causes *cannot* be invoked in the processes mentioned by Short. I shall first make the point that, no matter how different final causes are supposed to be from mechanical causes, if they are active then they must have effects. And to have effects in the world must make a difference to the world. To say this, I take it, is not to be involved in mechanistic bigotry. But now when we look at the details of the scientific account of the evolution of closed systems and of populations of organisms there is just no room, as it were, for these extra effects.

Consider an enclosed gas whose molecules are dispersed throughout its volume in a non-random way – for example, the molecules at one end of a box have, on average, a higher velocity than those at the other end. What will happen if no energy is allowed to enter from outside? If we assume a Laplacean deterministic model then the scientific account asserts, roughly, that the molecules will disperse throughout the volume in such a way that eventually the average velocities of any large selections of molecules, taken anywhere within the box, will be much the same. Fast molecules, through collisions, speed up slow ones and slow molecules similarly slow down fast ones. It is for this kind of reason that eventually, if the population of molecules is large, thermal equilibrium is reached.

Consider other enclosed volumes of gas, as many as you wish, whose (large) populations of molecules take up quite different non-random distributions. Precisely the same thing will happen in each of them (with very rare statistical exceptions). So all (or nearly all) of them will reach equilibrium for the same reasons involving mechanical interactions, etc. In each case the explanation for the final equilibrium state is a full one and does not mention any causes other than mechanical ones. So if the *only* causes acting in *each* isolated system that attains equilibrium are mechanical causes then the only causes acting in *all* the isolated systems must be mechanical causes. A general final cause does not get a look in and, in fact, if it did would, like Maxwell's demon, be expected to disturb the various states of equilibrium achieved in each system.

Macquarie University,
North Ryde, New South Wales 2113,
Australia

© ALAN OLDING 1985

ON TWO ARGUMENTS FOR COMPATIBILISM

By PETER VAN INWAGEN

THE two most popular arguments for the compatibility of free will and determinism are probably the following.

The *Ethics* Argument

> Analysis shows that statements of ability are disguised conditionals. More exactly, the correct analysis of 'X could have done A' is 'If X had decided (chosen willed . . .) to do A, X would have done A'. Therefore, having acted freely – having been able to act otherwise than one in fact did – is compatible with determinism (with the causal determination of one's acts).

The *Mind* Argument

> If one's acts were undetermined, they would be "bolts from the blue"; they would no more be *free* acts than they would if they

> had been caused by the manipulations of one's nervous system by a freakish demon. Therefore, free action is not merely *compatible* with determinism; it *entails* determinism.

(I call the first of these the *Ethics* argument because two classic statements of it are to be found in books of that title. See G. E. Moore, *Ethics* (London: Home University Library, 1912), Ch. VI; P. H. Nowell-Smith, *Ethics* (London: Penguin Books, 1954), Chs. 19 and 20. I call the second the *Mind* argument because it has appeared so often in the pages of that journal. See R. E. Hobart, 'Free Will as Involving Determination and Inconceivable Without It' (1934); P. H. Nowell-Smith, 'Free Will and Moral Responsibility' (1948); J. J. C. Smart, 'Free Will, Praise and Blame' (1961). A. J. Ayer's 'Freedom and Necessity', *Philosophical Essays* (London: Macmillan, 1954), did not appear in *Mind*. I am at a loss to account for this.) My statements of these two arguments are not intended to be complete or adequate. They are intended only to jog the reader's memory. For complete and adequate statements of the two arguments, the reader is referred to the places cited above.

I do not believe that it has been noticed that these two arguments are incompatible. That is: if either argument is sound (if its conclusion follows from its premises and if those premises are true), then the other is unsound.

To show this, we must first show that if the premise of the *Ethics* Argument is true, then free will is compatible not only with determinism but with *indeterminism* as well. To show this we need only tell a story having these three features: (1) in the story, an agent acts, and his act is causally undetermined; (2) in the story, his act is free, and (3) if statements of ability are disguised conditionals, then that story is internally consistent.

A story having these features is easy to construct. Suppose that Miss X has been deliberating about whether to tell the truth or to lie. Suppose that she has decided to tell the truth (and that, acting on this decision, she *has* told the truth). Suppose that this event – her having come to a decision to tell the truth – was not determined by earlier events or states: if God created a perfect duplicate of her as she was a moment before she made her decision to tell the truth, and if God placed that duplicate in circumstances identical with her circumstances at that moment, the duplicate might very well decide to lie. Suppose further that if X *had* chosen to lie, she would have lied. It is evident that (i) this story is consistent, and (ii) if 'X could have lied' is equivalent to 'if X had decided to lie, X would have lied', then the story entails the following three propositions:

- X told the truth
- X could have lied
- X's telling the truth was causally undetermined.

(I take the first two of these propositions jointly to entail 'X told the truth freely'.) It would seem that we have told a story with the required features, and have therefore demonstrated that if statements of ability are disguised conditionals, then free will is compatible with indeterminism. (It may be objected that there are available more sophisticated analyses of ability-statements as conditionals than the one that figures in the above argument. This is true, but I do not think that the argument would fail if it were reconstructed so as to involve any of these more sophisticated conditional analyses of ability. Of course, one could always offer *this* analysis: 'X could have done A' means 'If X had decided to do A, X would have done A, and either X's decision to do A or the non-occurrence of a decision by X to do A – whichever of the two in fact occurred – was causally determined'. But no one *has* ever offered any such analysis, and I can see no motivation for doing so, other than a desire to devise a version of the *Ethics* argument that is compatible with the *Mind* argument. But what could move anyone to devise a version of the *Ethics* argument that had *that* particular feature? A desire to have available as many arguments for compatibilism as possible?)

The promised conclusion is now easy to demonstrate. If the *Ethics* Argument is sound, then 'X could have done A' means 'If X had decided to do A, X would have done A'. If 'X could have done A' means 'If X had decided to do A, X would have done A', then (as we have seen) free will is compatible with indeterminism. If the *Mind* Argument is sound, then free will entails determinism. If free will entails determinism, then free will is not compatible with indeterminism. Therefore, if either argument is sound, the other is unsound.[1]

Syracuse University,
541 Hall of Languages,
Syracuse, New York 13210, U.S.A.

© PETER VAN INWAGEN 1985

---

[1] I wish to thank Mark Brown for suggesting an improvement in the structure of the argument, and Jonathan Bennett for the point about motivation in the second-to-last paragraph.

CAUSAL DEVIANCY AND MULTIPLE INTENTIONS: A REPLY TO JAMES MONTMARQUET

By JOHN BISHOP

ACCORDING to James Montmarquet, with a certain 'modest alteration', Christopher Peacocke's version of the causal theory of intentional action succeeds in solving the problem of causal deviancy ('Causal Deviancy and Multiple Intentions', ANALYSIS 42.2, March 1982, pp. 106–10). It is important to assess this claim, for, if it is correct, then the project of eliminating causal claims whose subjects belong in the category of substance in favour of those whose subjects belong in the category of events has finally accomplished its most formidable task, namely to make good this reduction with respect to attributions of intentional agency.

Montmarquet's revision of Peacocke's theory was partly prompted by a desire to defend it against earlier criticism of my own ('Peacocke on Intentional Action', ANALYSIS 41.2, March 1981, pp. 92–8). Rather than reassess what was perhaps an ill-stated objection, I shall hope to avoid tedium by considering on its own merits the account of intentional action which Montmarquet's discussion has yielded, an account which may be put as follows:

> S intentionally does X iff the causal chain connecting S's intention to do X with S's X-ing behaviour (a) meets Peacocke's strong DE requirement, and (b) has no proper part consisting of T's intention that S do X and S's X-ing behaviour which itself meets the DE requirement, where T is a distinct agent from S.

(See Montmarquet, p. 110. As actually stated, (b) excludes intervening intentions 'to do X' generally, but it is clear from the opening sentence that Montmarquet has in mind explicitly intentions *that S (the original agent)* do X. Peacocke's strong DE requirement is the requirement that S's intention to do X strongly differentially explain the behaviour which matches it. This means that the causal laws of the explanation must specify S's X-ing as a function of S's intention (under their relevant physical descriptions) in a way which satisfies the condition of 'step-wise recoverability'. See Peacocke's *Holistic Explanation*, Oxford, 1979 for the details, especially pp. 66 and 80.)

It is an apparent implication of this – let us call it the P_m – account that if S successfully carried out a deliberate plan to make a certain event, *e*, occur by getting another agent, T, intentionally to produce *e*, then S does not himself intentionally bring *e* about (though T may do so). Or, at least, this is implied if part of what T intends in acting is that S should perform the intentional act of making *e* happen – and surely T will intend this if he seeks to act on S's behalf or under S's command? Montmarquet steers round this unwelcome consequence by distinguishing an intention *to do X (oneself)* from an intention *to have X done (by someone else for one)*, and, though this is not made explicit, the P_m account is meant to apply only when S's intention to do X does *not* have the latter form. It is preferable, I suggest, to construe P_m as proposing sufficient and necessary conditions of *basic* intentional action only, which is, after all, what Peacocke's original condition was intended to provide (Peacocke, p. 88). Succeeding in carrying out an intention that someone else do something for one is not a less than 'full-fledged' intentional action, as Montmarquet implies (p. 109): it is simply a species of *non-basic* action – one which can be achieved only by first executing other more immediately achievable intentions.

To understand the point of difference between P_m and Peacocke's own account of basic intentional action, we need to consider their respective responses to the powerful neurophysiologist example. Peacocke allows that the DE requirement may still be satisfied when an agent's body is moved by the intentional action of a second agent in accordance with his neurophysiologically-based knowledge of the first agent's basic intentions. Peacocke accordingly augments his DE requirement with the condition that 'the chain from intention to bodily movement should not run through the intentions of another person' (p. 88). But Montmarquet argues that this condition is too strong: not all such chains need be counted deviant. To show this, Montmarquet adapts an already abstruse example – the lightning-assisted gunman case originally due to Pears (Montmarquet, p. 107; Peacocke, p. 93). We are to consider an agent, G, whose efferent nerve to his index finger is severed, but whose brain is so 'wired up' to that of a second agent, H, that when G initiates the action of moving his index finger, H is caused to form the intention that G should so move his (G's) finger, and this intention – it so happens! – causes the required electro-chemical impulse 'in the severed section of G's nerve' (see Montmarquet, p. 108). In this case, Montmarquet believes, G does intentionally move his finger, even though the relevant causal chain passes through another's intentions.

This proposed counterexample stretches empirical credulity so far that its claim to engage useful intuitions must surely be suspect. Indeed, the detail of the case is puzzling. In Pears' original example, a flash of lightning carries the impulse across the severed nerve – a phenomenon just within the realm of conceivable science-fiction. But how can 'wiring' two *brains* together restore electro-chemical communication between the severed ends of an *efferent* nerve, as is envisaged in the present example? Besides, Montmarquet seems confused about how to judge this case. He introduces it as an, albeit highly unusual, case of non-deviant basic intentional action. Yet later on he says that the causal chain connecting 'the second (viz., H's) intention with G's subsequent behaviour' will be 'classified as deviant by its evident inability to meet (Peacocke's) *un*augmented DE requirement' (his emphasis, p. 109), in which case DE can hardly be satisfied for the whole chain from G's intention to G's matching behaviour, and so, by Montmarquet's own theory (P_m), this case must be deviant.

Nevertheless, however unhelpful Montmarquet's attempt at a counterexample may be, his objection that Peacocke's augmentation 'makes the resulting condition so strong that it no longer states a necessary condition of intentional action' (p. 107) is surely sustainable. Physical agents exercise their control in a repertoire of basic actions only by dint of appropriate physical mechanisms.

Physiology studies the standard mechanisms involved; but there may also be non-standard ones, as reflection on the history of surgical prosthetics will convince us. It is strictly possible, moreover, that *any* portion of a physiological mechanism should be replaced by a prosthesis which preserves its function, even if the mechanism which goes proxy operates on a unique occasion as (so to say) a purely fortuitous prosthesis. Now, I take it that Montmarquet's example is meant to illustrate the abstract possibility that *another agent's intentions should form part of the prosthetic mechanism* through which, on perhaps only a single occasion, an agent exercises his control over his bodily movements. A neurological prosthesis for basic actions of a given type is a strict possibility – and, if flashes of lightning can without absurdity be imagined to function fortuitously as part of such a prosthesis, why couldn't *some* arrangement involving another agent's intentions function in the same way?

It turns out, however, that if these are good reasons for rejecting Peacocke's augmentation to DE, they are also good reasons for rejecting Montmarquet's alternative (part (b) of P_m), the spirit of which is to suggest that, while the mere occurrence of T's intention that S do X in the causal chain from S's intention to do X to S's X-ing behaviour does not of itself introduce deviance, if that portion of the causal chain initiated by T's intention constitutes *T's* intentional *action* of bringing it about that S does X, then the case is deviant. (I cannot forbear the quibble that the letter of clause (b) in P_m does not quite capture the spirit of this suggestion. For, as it stands, it excludes not only proper parts of the causal chain which count as T's intentionally making S do X but also those which, though they meet DE, nevertheless count as deviant semblances of this action. If Montmarquet is not content to have all such cases excluded, then the point is more than a quibble, since clause (b) will have to be supplemented with whatever extra condition it is which ensures that the part of the chain which satisfies DE also counts as T's intentional action.)

But why should Montmarquet's proposed addition to the bare DE requirement count as an improvement, given that the essential flaw with Peacocke's was his overlooking (in effect) the possibility that a neurological prosthesis for a basic action should include another agent's intentions as part of its causal mechanism? For surely it is purely arbitrary to allow that a prosthetic mechanism can include another's intentions but not his intentional actions – at least not if these are describable as his intentionally producing precisely the behaviour which the original agent intended? It seems absurd to maintain that, while other minds can conceivably supply aid to an agent whose neural mechanisms are impaired, they can do so only if they are blind to this service – only if, that is, they don't *intentionally* produce just the very state of his body which the impaired agent intends.

But if we allow that another agent's intentional action in making S do X can be part of a prosthetic aid which allows *S* himself to carry out his basic intention to do X, do we not then vindicate Peacocke's unaugmented DE requirement? Does it not emerge that, provided DE obtains, a causal chain may pass through the intentions or the intentional actions of other agents without acquiring the least deviance? I think not. For what I have argued is just that T's intentional action in bringing it about that S does X *may* be part of a prosthetic mechanism which permits S to perform the action of doing X. But not all interventions of the intentional activity of a second agent qualify as *part of the mechanism* through which the agent non-standardly exercises his own control. Some such interventions, indeed, pre-empt the original agent's very opportunity to carry out his basic intention and so preclude his behaviour from counting as *his* intentional basic action. For an agent performs a basic intentional action only if his forming (or having) the relevant basic intention still *leaves it up to him* to carry out that intention or not. And, where pre-emption occurs, it is no longer up to the agent whether his basic intention is matched. (Of course, it is true that once an agent has a genuinely basic intention it is not up to him to *deliberate* as to how to achieve it – but this does not entail that it is not up to him to carry it out.)

Thus, our response to the powerful neurophysiologist example depends crucially on whether we construe the intervention as pre-emptive or not. If, as Peacocke's wording suggests, intervention is triggered immediately the agent's intention is formed (let us grant that detection of it is instantaneous), then the neurophysiologist does not leave it up to the agent to carry out his intention and, even if DE is satisfied as it surely could be, we will not have a case of non-deviant basic intentional action on the part of the original agent. On the other hand, if we redescribe the case to engage the intuition that the agent does at least begin the process of carrying out his intention, then the neurophysiologist's intervention may appear simply as part of a non-standard mechanism for the exercise of the agent's own control. Thus, with Montmarquet's case (4), where the neurophysiologist is described as acting upon his recognizing the agent's 'intention *and the cortical event initiating his action*' (my emphasis, p. 108), it seems to me that the agent *does* carry out his intention, though only by grace of a physical set-up which includes the neurophysiologist's intentional act. What is special here is not that the agent has to rely on the operation of a system beyond his control in order to perform even a basic action of his own – for that applies to *all* the acts of physically embodied agents. What is special in such a case is just that the physical mechanism which secures 'the medium' of his control is non-standard. Not all non-standard arrangements would be fit vehicles for an agent's exercise of control, of course – but it does seem a fair hypothesis that those which satisfy some 'responsiveness' condition (i.e. of the matching behaviour to the content of the agent's intention), such as Peacocke's DE requirement, may properly be counted so – always provided that there is no pre-emption of the agent's opportunity to carry out his basic intention.

Precise physical criteria for determining whether another agent's intervention is pre-emptive or not are no doubt extremely difficult to provide. Indeed, I am prepared to wager that if we ever encounter any actual instances of 'multiple intention' basic intentional acts of the sort discussed here, we shall be able to describe them to fit either conviction. What is important, however, is that we bring to our assessment of any physically described situation the *concept* of a distinction between pre-emptive and non-pre-emptive intervention by other agents. And the concept of this distinction is quite independent of the concept of satisfaction of the strong DE requirement, since a chain which satisfies this requirement and which includes a second agent's intentional action may or may not be pre-emptive.

Montmarquet does indeed recognize the distinction between 'pre-empting the agent's opportunity to act by intervening before the agent is able to initiate any action at all' and intervening so as to 'complete an action which (the agent) has already begun for himself' (p. 108). But he regards this distinction as non-crucial, and instead places emphasis on the question whether or not the second agent's intention issues in an action of his. My submission, in sum, has been that Montmarquet has gone for the wrong distinction, and, in consequence, has persuaded himself of the importantly mistaken view that considerations about the 'sensitivity' of matching behaviour to the intention which caused it are alone sufficient to provide conditions which exclude deviance. In fact, the truth is that our actual concept of what it is for an agent to perform an intentional action includes the notion that the agent must himself be the originator of the action. And what it means to be an originator of a causal chain which constitutes an action cannot be reduced to a matter of the chain's satisfying some responsiveness condition such as the DE requirement. That we can make sense (within the categories of our 'folk psychology' of action) of a case where responsiveness is satisfied but the agent whose intention is concerned nevertheless does not even have the chance to originate the action is enough to establish the point. But the origination condition is not only irreducible to any responsiveness condition – it is also rather more subtle than might have been expected, for its satisfaction does not exclude all forms of intervention by the intentional activity of others. Even basic actions could in theory be carried out by dint of non-standard mechanisms which include 'multiple intentions', and, indeed, multiple intentional actions.

The University of Auckland,
Auckland, New Zealand

© John Bishop 1985

LYONS AND TIGERS

By PAUL SIMPSON

THE conclusion reached by William Lyons, 'The Tiger and his Stripes' (ANALYSIS 44.2, March 1984), is up to a point acceptable. It is: 'one can count stripes on imagined tigers but one must do so in the same way as one counts on real tigers'. What is not acceptable is the method of counting he proposes: 'a repeat in imagination of what we do in perception, that is, engage in a long series of "snapshots" of single, fingered stripes'. How does one put one's *finger* on the stripes of an image?

His answer: 'perhaps with our finger *pictured* as picking out the stripes' (my emphasis). When put to the test, which I have put to several people, the addition of a human finger to the image of a tiger – or to a series of images – only serves to distract from the set task of counting stripes. One cannot run one's flesh and blood finger along the hides of imagined tigers; and its imaged counterpart is useless as a counting aid.

The counting of stripes of images is best accomplished by attaining the skills needed in counting elements within holistic percepts, skills that Lyons denies to perceivers of both real and imagined objects. These skills now have minimal application in trade and technology. They are hence no longer developed educationally or professionally. They were once widely possessed by craftsmen, in particular those who practised the hand-setting of individual letters and spaces of lines of type – so that they could calculate at a glance the number of insertions needed to rectify right-hand margins. A related skill is still retained by some knitters, and to varying extents by those who can translate musical notation to the keyboard by sight-reading; but the musicians usually pause before being prepared to verbalize the number of notes being conveyed to their hand-movements after their swift inspection of a given set of bars.

My quarrel is not with Lyons' initial criticism of Daniel Dennett's 'put-down' of images on the curious charge that imagined tigers have uncountable stripes, if any, and hence do not duplicate the features of real tigers (D. C. Dennett, *Content and Consciousness*, Routledge, 1969, pp. 136–7). There is no application here of the principle of the indiscernibility of identicals, or the law of the substitutivity of identity, as no one claims that imaged tigers are identical with real tigers. However, as we shall see, some people experience images that are, as far as we can tell, faithful replicas of original appearances. My quarrel with Lyons starts when he offers correctives to Alastair Hannay's and J. A. Fodor's defences of images (A. Hannay, *Mental Images, a Defence*, Allen & Unwin 1971, V.3; J. A. Fodor, *The Language of Thought*, Crowell 1975, pp. 187–91) – whose features they hold to be uncountable owing to the vagueness, blurriness and evanescence they find inherent in images. Lyons concedes in one way too much and in another way too little to Hannay and Fodor, when he proceeds to transpose the difficulties of counting accurately the stripes of real tigers to the counting of stripes of imaged tigers. He concedes too much in proposing difficulties, that do *not* stand in the way of skilled counters and estimators of quantities within holistic percepts. He concedes too little in restricting the images with which he is concerned to clear, detailed, stable, and evidently repeatable representations.

Standard reference works in empirical psychology reveal a continuum of human capabilities in set tasks of visual image formation when measures are applied to the images' accuracy, detail, retention time, stability and reproducibility *ad lib.* At one extreme there are those (Ryle? Sartre?) whose accounts indicate an incapability for conscious visualization – a condition which is neither rare nor incapacitating. There are some suggestions that the condition is often associated with high levels of linguistic competence and associated too with characteristic EEG readings of the alpha rhythms of adult brains. The absence of alpha rhythms in adults has been regarded as a sign of vivid visual imagination. At an extreme of several scales on which 'quality' can be ordered are the images of so-called eiditikers (William Blake). Contrary to popular mythology, research findings in this area are more compelling than mere speculation based on subjective reportage from experimental subjects. Especially when appropriate (eidetic) subjects are asked to recapture previously witnessed appearances, quality can be independently checked. Thus we find the following results:

> Haber [J. Leask, R. N. Haber, R. B. Haber, 'Eidetic imagery in children; II: Longitudinal and experimental results', unpublished 1968] gave his eidetic subjects a puzzle picture in which two schematic drawings of ocean scenes revealed a bearded man when superimposed; a few of his eidetic subjects were able to superimpose eidetic images of the two ocean scenes, and reported how the face sprung into view. Stromeyer and Psotka [C. F. Stromeyer and J. Psotka, 'The detailed texture of eidetic images', *Nature* 1970, pp. 346–9] ... used Julesz patterns to test a young woman who possesses unusually vivid eidetic imagery. She was asked to recall via eidetic imagery a computer-generated pattern presented to her left eye several hours or days earlier while its matched pair was [at the later time] presented to her right eye. The central three-dimensional square, reported by subjects who view the two patterns simultaneously, was clearly visible to this young woman. (S. J. Segal, 'Processing of the Stimulus in Imagery and Perception', in S. J. Segal, ed., *Imagery, Current Cognitive Approaches*, Academic Press, N.Y., 1971, p. 73.)

In order to count accurately the stripes of an imaged tiger, two capacities are needed. The first is to be granted a sufficiently high quality of one's image. It needs at least to be stable, but not necessarily to the above extremes. Secondly, in estimating quantities, it will be necessary to acquire a skill that modern technology has rendered redundant. To many trades it was important that these estimates be accurate. They were not gained by guesswork, although they do appear to have relied upon algorithms, groupings, multiplication (as in finding the number of tiles on a roof top), etc. Perhaps Lyons would not extend the description 'counting' to these procedures, reserving counting for one-to-one finger-guided mappings onto the natural numbers; but his text allows that we can, if we please, 'take' the stripes in 'twos, threes, etc.'. The issue is of no importance here, as Dennett's original argument was meant to depend on the indeterminacy of the *number* of the imaged tiger's stripes, not the method of that number's assessment.

The method of the number's assessment recommended by Lyons requires not only the mysterious exercise of a pictured finger but a skill in gaining images that are 'blow-ups' of details in a composite image, in order to form a temporally ordered but complete series.

> Since imagination is parasitic on perception of objects in public space, then the imagined [sic] counting of stripes can at most be a repeat in imagination of what we do in perception, that is, engage in a long series of 'snapshots' of single, fingered stripes.

He should try mounting a real but single snapshot of a tiger on a wall at a distance beyond the reach of his fingers. He will find, after some practice, that he can count the visible stripes that are represented on the picture's surface. A person whose visual images are appropriately detailed and stable can perform the same feat after removal of the snapshot, even if not pre-warned that information will be sought about the stripes. The accuracy of the answer is similarly checkable upon someone else's examination of the hidden original, but the second person may make use of the fingering technique if that helps. The first would be unable to, and unwise to try.

University of Tasmania,
Box 252C, G.P.O. Hobart,
Tasmania, Australia 7001

© PAUL SIMPSON 1985

HARE ON MORAL CONFLICTS

By IGOR PRIMORAC

PROFESSOR Hare's two-level theory of moral thinking (R. M. Hare, *Moral Thinking, its Levels, Method and Point*, Oxford 1981) is meant to bridge the gap between intuitionism and utilitarianism by integrating whatever is true in the former into a sophisticated version of the latter. The synthesis thus achieved is then used to deal with a number of problems in normative ethics; in particular, it provides what Hare believes to be a rational and workable procedure for resolving moral conflicts.

The difficulties all the main ethical theories have with moral conflicts are well known. Intuitionism accounts for the role moral rules play in moral reasoning, but presents them as a plurality of intuitions which in principle cannot be organized into a hierarchy, and therefore remains helpless when they conflict with one another. Act-utilitarianism offers an efficient method for solving moral conflicts, but at the price of making us forget about the rules and embrace all sorts of unacceptably permissive conclusions. Rule-utilitarianism attempts to avoid this permissiveness by bringing moral rules back into the picture; it gives them a utilitarian justification, which is held to be a distinct advantage over intuitionism. However, when we take a closer look at the options available to the rule-utilitarian in a case which calls for an exception to a moral rule because its utilitarian justification does not apply, it turns out that we are either led all the way back to act-utilitarian permissiveness, or called upon to engage in rule-worship.[1]

Hare's point of departure is the claim that moral philosophy must take into account the very human limitations of the moral agent. Our knowledge is limited, and so are our capacities for clear thinking; we tend to be partial; on top of that, we often have to come to a decision within a very short time. Therefore it will not be for the best to keep questioning such simple, basic moral rules as those that prohibit breaking promises, lying, or stealing, and to make our decisions by balancing the good and bad consequences of options facing us in each particular case, the way act-utiliarians want us to do. The utilitarian point of morality is much more likely to be promoted if we comply with these rules – not always, to be sure, but most of the time. This, then, is what we ought to do; in order to be able to do that, we should develop strong dispositions against breaking such rules, and come to experience feelings of guilt whenever we do break any of them.

Hare calls the level at which we recognize and apply such rules the intuitive level of moral thinking, partly because it is the intuitionist who insists on their importance. As against the latter, he points out the need for such rules to be justified, and to be qualified from time to time in view of cases which they do not fit; he also points out the need for a procedure for dealing with situations in which such rules come into conflict with one another. The selection and revision of moral rules, and the adjudication of conflicts between them, is done at another, critical level of moral thinking. It will be safe to launch into this kind of thinking only at times when we are free from the limitations typical of our everyday moral life, for it is purely utilitarian: it selects and revises moral rules to be used at the intuitive level in terms of their acceptance-utility, and resolves conflicts between them by weighing and balancing consequences of options in particular situations. If we were completely impartial, omniscient, and had superhuman powers of clear, logical thinking, we could safely deal with each particular case of moral choice in this way, and be certain that we should always do what *is* right. As we are only human, most of the time we ought to keep to the intuitive level of moral thinking, go by the rules, and thus do what is *most likely* to be right.

---

[1] The point concerning rule-utilitarianism might be thought controversial; I have tried to establish it on pp. 595-613 of my 'Utilitarianism and Punishment of the Innocent', *Rivista Internazionale di Filosofia del Diritto*, vol. 57 (1980).

Actually, we shall not leave the intuitive level and take to critical thinking in all cases of moral conflict. Sometimes we shall just 'feel sure' that one of the basic moral rules overrides the other in the particular case at hand. In other cases we shall set aside the conflicting rules and consider the particular situation directly, bringing critical thinking to bear on it and deciding for the alternative with better consequences.[2] Whichever way we resolve the conflict, we shall also feel guilty for having offended against a moral rule. This will indicate that we have adopted the rule and are likely to offend against it only rarely and with considerable reluctance, so there are good utilitarian reasons for developing and cultivating this feeling. Thus the unacceptable permissiveness that plagues both actutilitarianism and some rule-utilitarian views seems to have been avoided; because of practical limitations inherent in everyday life, we ought to keep to the rules on most occasions. But we are not invited to take to rule-worship, to which some rule-utilitarians resort: there will be times when we shall go against a moral rule.

Is this, then, a plausible account of moral conflict? It will come natural to go by the basic, everyday rules most of the time, as Hare wants us to; difficulties start when two such rules come into conflict. If these rules have been selected properly, such conflicts will be an exception rather than the rule. But can this really be ensured? When giving examples of the kind of rules he has in mind for the intuitive level, Hare mentions those that prohibit breaking promises, lying, stealing, violence, murder – and also the rule of beneficence (ibid. p. 121; see also pp. 117-18, 193). Now if beneficence is one of the basic moral rules, on an equal footing with the rule of promise-keeping, for instance, then each and every case in which to keep a promise is not at the same time the course of action with the best consequences attainable under the circumstances will be one of conflict of the two rules (cf. J. J. C. Smart, 'Extreme and Restricted Utilitarianism', *Philosophical Quarterly* 1956, p. 354). We would be able most of the time simply to keep to the rule of making good our promises, as Hare suggests, if all other basic rules were similarly specific, for most of the time this rule does not conflict with those others – and when it does, it is clear that that is the case. Then the usual situation of choice involving a promise would be one to which only one moral rule – that of promise-keeping – is relevant. Once the rule of beneficence is brought into the picture at the same level, however, on each and every occasion on which it is not quite clear that to keep the promise is the option with the best consequences attainable under the circumstances – and this is often not quite clear, which is just the point of having the rule – we shall have to take up the question whether we are not faced with a conflict of the two rules. Thus the rule of promise-keeping becomes liable to be challenged much more frequently than Hare expects.

---

[2] The question just when we should deal with moral conflict on the intuitive level, and when should we engage in critical thinking, does not admit of a philosophical answer; 'it depends on what powers of thought and character each one of us, for the time being, thinks he possesses' (R. M. Hare, op. cit. p. 45).

In some cases we shall resolve the question intuitively, without resorting to critical thinking. On the *decisions* reached in this way nothing can be said in general; to the *way* of reaching them, though, one could apply Hare's own comment on the intuitionist's account: since we are not told just *how* to assess the comparative moral weight of the rules, this is 'less a method than an evasion of the problem' (op. cit. p. 34). On other occasions we shall apply critical thinking – we shall weigh and balance the good and bad consequences of the particular alternatives, as act-utilitarians do. Whenever it turns out that to keep the promise is actually the most beneficent option – i.e. that there is no conflict of rules after all – we shall decide to keep it. Whenever it turns out that there *is* a conflict of the two rules, on the other hand, we shall decide not to keep it. Since the rule of beneficence is present at both levels, first as one of the two conflicting first-order rules, and then again as the principle of critical thinking for adjudicating conflicts of such rules – since it is both a litigant and the judge – it is certain to prevail in each and every case of this sort. In all such cases, that is, we shall end up with the old act-utilitarian permissiveness on breaking promises, lying, stealing, engaging in violence and so forth.

Nor can we really be expected to feel guilty about it. We shall know that it would be good for us to feel guilty, for reasons which Hare's two-level utilitarianism adduces – but we shall also know that, from the point of view of the same theory, we *are not* guilty. If Hare nevertheless manages to get us to feel guilty, although we are innocent, this will be a rather compromising achievement; for, as Walzer has pointed out, it will be 'another case where suffering of the innocent is permitted and even encouraged by utilitarian calculations' (M. Walzer, 'Political Action, the Problem of Dirty Hands', *Philosophy and Public Affairs* 1972–3, p. 173).

In the present context, though, the question of guilt is secondary. The main point is this unacceptably far-reaching scope of the rule of beneficence which, against Hare's intentions, gives the two-level theory a strong push in the direction of act-utilitarian permissiveness. Now Hare does not attempt anything like a list of simple, everyday rules that make up the intuitive level of moral thinking; he only gives a few examples of such rules. Could it be that the rule of beneficence is merely an infelicitous illustration – that the problem will go away if we just forget about beneficence as a first-level rule? Things are not that simple, however, for it was to be expected that some such rule would be found at this level. If the point of morality as a whole is doing good in the world, as it is for Hare (and any utilitarian), then there has to be some rule to that effect, alongside all the rules on promise-keeping, truthfulness, abstaining from violence and the like. For these latter, more specific rules can never between them effectively encompass the whole field of possible beneficence.

Could the problem be solved, perhaps, by giving the rule of beneficence a special, residual status, making it relevant only in those cases to which no other, more specific rule applies?[3] This will not do either, for it would mean introducing a hierarchy within moral rules at the first, intuitive level. And Hare expressly rejects the idea of such a hierarchy, not only for practical reasons (which are not relevant here, for the proposed hierarchy would be quite simple and easy to handle), but also because 'we are unlikely to be able always to put principles in the same order of priority' (loc. cit.). It would be quite implausible to maintain that the rule of promise-keeping, for instance, must *always* come before that of beneficence. (To be sure, Hare allows for sub-levels within the first, intuitive level of moral thinking, but he does that in the context of a discussion of supererogation, and the differentiation is in terms of scope of the rules – those that apply to everyone, those pertaining to specific social roles, and those very personal ones – not in terms of their comparative stringency; op. cit. pp. 199–203.)

Hare criticizes intuitionism for being indefinite and unhelpful with moral conflicts. He rejects act-utilitarianism, partly on account of its permissiveness with respect to simple, everyday moral rules. His two-level theory of moral thinking is advanced as a synthetic one, which avoids both these faults, while appropriating the truth contained in both theories. But it seems to be vitiated by a combination of the very defects it set out to avoid.[4]

The Hebrew University,
Jerusalem, Israel

© IGOR PRIMORAC 1985

---

[3] This move was suggested to me as a plausible defence of Hare's theory by Prof. Ruth Gavison.

[4] Thanks are due to a referee for *Analysis*, whose criticism has forced me to reformulate a crucial step in the argument.

A NOTE ON SUBSTANCE CONCEPTS

By B. J. GARRETT

IN *Sameness and Substance* (Oxford, 1980) David Wiggins argues for a distinction, within the class of sortal concepts, between 'phased concepts' and 'substance concepts'. He provides the following two characterizations of the notion of a substance concept. According to the first characterization, substance concepts are 'sortal concepts that present-tensedly apply to an individual x at every moment throughout x's existence, e.g. *human being*...' (p. 24). According to the second characterization if '... "x is no longer f" entails "x is no longer", the concept that the predicate stands for is in my usage a *substance concept*' (p. 64).

These characterizations, however, are not equivalent. The concept *person*, for example, qualifies as a substance concept according to the second characterization, but not according to the first. For it is plausible to suppose that if x is a person then x could not cease to be a person without ceasing to be. Yet, assuming that zygotes or foetuses are not persons, there was a time at which x existed and was not a person. Thus the two characterizations are not equivalent; though it is clear, I think, that it is the first characterization which Wiggins requires for his theory of individuation.

Moreover the second characterization would have the consequence that, for example, the concept *person born in Edinburgh* would qualify as a substance concept. And this would sever the connection between the notion of a substance concept and the notion of a *fundamental* sort of thing. However, this problem could be circumvented by stipulating that only 'unrestricted' sortal concepts can qualify as substance concepts. Indeed such a qualification is going to be required for the first characterization also, otherwise the concept *human being who will live for 30 years* would qualify as a substance concept. (Actually I disagree with Wiggins that the unrestricted sortal concept *human being* is a substance concept. For I can see no *a priori* reason why a person who is a human being could not, with the aid of some fantastic body-transfer machine, come to occupy a non-human, or even non-animal, body and thus cease to be a human being. If this is so then, on either characterization, *human being* is not a substance concept.)

Note finally that neither of the two characterizations carries any essentialist import. The first characterization clearly does not; and nor does the second: it may be true that x cannot cease to be a person without ceasing to be, yet *being a person* is not an essential property of x since there was a time, in the actual world, at which x existed and was not a person.

Lincoln College, Oxford

© B. J. GARRETT 1985

INQUIRY

Editor: Alastair Hannay

Vol. 27 (1984), No. 4	
Pragmatics in Science and Theory in Common Sense	K.V. Wilkes
Anti-reductionist Materialism	Kathleen Lennon
Consciousness and Content-formation	Philip Cam
Sperry's Concept of Consciousness	Charles Ripley
The Agent in a Northern Landscape	David Londey
Non-human Rights: An Idealist Perspective	T.L.S. Sprigge
A Case for Idealism? (John Foster, *The Case for Idealism)*	Frank Jackson
Searle on Intentionality (John R. Searle, *Intentionality: An Essay in the Philosophy of Mind)*	Ronald McIntyre
Vol. 28 (1985), No. 1	
The Role of Rules in Ethical Decision Making	Eugene C. Hargrove
Speech and the Social Contract	Roy Turner
Rationality, Democracy, and Freedom in Marxist Critiques of Hegel's Philosophy of Right	David Campbell
Objectless Activity: Marx's 'Theses on Feuerbach'	A. Giles-Peters
Performative, Subject, and Proposition in Habermas's Theory of Communication	Erling Skjei
Reply to Skjei	Jürgen Habermas
Comment on Steven Yates's 'Feyerabend's Democratic Relativism'	C. Fred Alford
Understanding Peter Winch	R.J. Anderson and W.W. Sharrock
Social Science or Dialogues of the Deaf (W.G. Runciman, *Treatise on Social Theory,* I)	Michael O. Luntley

Subscription price:
Institutions: N.kr. 300,- $50.00 — Individuals N.kr. 155,- $26.00.

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, 0608 Oslo 6, Norway — P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

MIND

Published for the Mind Association
by Oxford University Press

From 1985 (Volume 94) Mind will have a new Editor, **Simon Blackburn** of Pembroke College, Oxford, and a new Publisher, **Oxford University Press**

Forthcoming papers include:

Ralph Walker on Spinoza and the Coherence Theory of truth, **Gilbert Plumer** on the Specious Present, **Peter Forrest** on Backward Causation and Free-will, **Charles Travis** on Vagueness, **Frank Jackson** on Semantics and Deontic Logic, **Philippa Foot** on Benevolence and Consequentialism

1985 rates:
Personal Subscribers: £10 UK, $22.50 N. America, £13.50 Elsewhere
Institutional Subscribers: £16 UK, $32.50 N. America £19 ROW

ORDER FORM

☐ I wish to subscribe to Mind Vol. 94

☐ Please send me an inspection copy

Payment may be made by
Visa or Access Card

Name ....

Address ....

....

Postcode .... Country ....

Oxford Journals

Oxford University Press, Walton Street, Oxford OX2 6DP, UK

Thinking about Social Thinking

An Introduction to the Philosophy of the Social Sciences

ANTONY FLEW

Antony Flew insists throughout this provocative book that the social sciences have attracted more than their fair share of misconceptions and 'wishful thinking', and attempts to right these wrongs by giving the reader a clear view of the goals and methods of modern social science.

200 pages, hardback **£16.50** (0 631 14189 8)
paperback **£5.95** (0 631 14191 X)

Contemporary Epistemology

An Introduction

JONATHAN DANCY

'Worthwhile and fresh. The book conveys a strong and deserved impression of good sense, and assists the reader (even the inexperienced reader) in sorting out arcane arguments and distinctions as if they were relatively familiar – as indeed they become, with Dancy's help.' Joseph Margolis.

272 pages, hardback **£22.50** (0 631 13621 5)
paperback **£7.50** (0 631 13622 3)

Needs, Values, Deliberation and Truth

Essays in the Philosophy of Value

DAVID WIGGINS

This collection comprises revised or corrected versions of a number of essays, as well as several important new ones. An introduction draws together the recurrent themes, most notably objectivity, subjectivity, intersubjectivity, and the scope and limits of the aspiration to find plain truth in the judgements of morals, politics and aethetics.

240 pages, **£12.50** (0 631 14044 1)

Wittgenstein and Derrida

HENRY STATEN

Although Derrida has had a great impact on literary theory, the philosophical foundations of his writings have not been well understood, and his early work has received little attention. Here Henry Staten examines the relation of his ideas to the classical philosophical concepts, and to the later work of Wittgenstein.

224 pages, **£19.50** (0 631 13982 6)

Personal Being

ROM HARRÉ

'Brilliant and endlessly provoking . . . *Personal Being* provides a radical new schema by which we may align future work in the psychology of intentionality, rationality and emotion, and meaning.' *Times Literary Supplement*

'Distinguished by imaginative scope, richness of argument and penetrating criticism of alternative approaches.' *British Book News*

310 pages, hardback **£17.00** (0 631 13318 6)
paperback **£5.95** (0 631 14228 2)

Basil Blackwell

Understanding Social Science

A Philosophical Introduction to the Social Sciences
ROGER TRIGG

This introductory textbook for anyone studying the nature of social science questions its scientific basis and shows that philosophical presuppositions are a necessary starting point for any study of society.
220 pages, hardback **£16.50** (0 631 13365 8)
paperback **£5.95** (0 631 14161 8)

Reality and Reason

SEAN SAYERS

Sean Sayers develops and defends a dialectical and realist theory of knowledge covering, as he does so, a number of fundamental areas in epistemology. His basic thesis is that all ideas reflect reality, and this he elaborates with reference to analytical and continental philosophy, and to Descartes, Locke, Berkeley, Hume, Kant and Hegel.
272 pages, hardback **£17.50** (0 631 14022 0)
paperback **£5.95** (0 631 14274 6)

The State

ANTHONY DE JASAY

This book traces the logical and historical progression of the state from minimalist protector of life and property, through agile seducer of democratic majorities, to the welfare-dispensing drudge that it is in many countries today. How can it escape from the treadmill? Is the rational next step a totalitarian enhancement of its powers? The author presents a disturbingly logical 'agenda' for the state in pursuit of its 'self-fulfilment'.
304 pages, **£17.50** (0 631 14025 5)

Morality and Conflict

STUART HAMPSHIRE

'Characteristically authoritative . . . Stuart Hampshire compels us to think. A single review cannot do justice to its richness, nor indeed to the moving seriousness with which it is written.' *The Listener*
'Important and influential essays' *Times Higher Education Supplement*
186 pages, hardback **£16.50** (0 631 13336 4)
paperback **£5.50** (0 631 13982 X)

Kant's Political Philosophy

HOWARD L. WILLIAMS

'A welcome addition to the small amount of literature devoted to this aspect of Kant work . . . a clear and useful outline of all the major themes . . . will be helpful to undergraduates coming to grips for the first time with Kant's very difficult work . . .'
S.P. Sayers, *British Book News*
304 pages, hardback **£20.00** (0 631 13123 X)
paperback **£7.50** (0 631 14153 7)

Basil Blackwell

Philosophical Investigations

Edited by **D.Z. Phillips**

Volume 8 No. 3 July 1985

O. K. BOUWSMA: PHILOSOPHY AND TEACHER

Articles:

Remarks on Bouwsma's Method of Failure — J. L. CRAFT
'There Is In Wittgenstein's Work No Argument And No Conclusion' — MICHAEL E. MALONE
Bouwsma on Moore's Proof — JAMES D. CARNEY
Sense and Sensibility — LARRY KIMMEL
O. K. Bouwsma, 1937 and Later — C. D. ROLLINS

Reviews:

J. L. Craft and Ronald E. Hustwit (eds.), O. K. Bouwsma, *Towards a New Sensibility* — COLIN LYAS
J. L. Craft and Ronald E. Hustwit (eds.), O. K. Bouwsma, *Without Proof or Evidence* — WALFORD GEALY

Books Received

Published quarterly: January, April, July and October
Subscriptions to Volume 8, 1985
Institutions: *£28.00 (UK) £35.00 (Overseas) $65.55 (US) $80.30 (Canada)*
Individuals: *£12.95 (UK) £16.20 (Overseas) $30.35 (US) $37.20 (Canada)*

Orders, with remittance, or sample copy requests, should be sent to:
Iris Taylor, Journals Department, Basil Blackwell,
108 Cowley Road, Oxford OX4 1JF, England

Basil Blackwell

NOTES

The ANALYSIS Committee consists of: Chairman, P. T. Geach; Secretary, T. R. Baldwin; Andrew Harrison, R. H. Holland; J. H. McDowell, D. H. Mellor, M. A. Stewart, R. G. Swinburne, A. R. White, C. J. F. Williams, Peter Winch. This committee is responsible for appointing and advising the Editor and for the general policy of the paper.

SUBSCRIPTIONS. ANALYSIS is published by Basil Blackwell Publisher Ltd, 108 Cowley Road, Oxford OX4 1JF, England. Each volume comprises four numbers, each of at least 48 pages, appearing in January, March, June and October. Orders should be sent to the publisher. Subscriptions prices are as follows:

25 MAY 1989

	U.K.	Overseas	U.S.A.	Canada
Institutions	£11.00	£13.50	$33.00	$39.50
Individuals	£7.45	£9.25	$21.00	$25.00

BACK ISSUES. Volumes 1–37 are now available in complete volumes in an offset reprint from Swetz & Zeitlinger, 471 Keizersgracht, Amsterdam, Holland. Recent back issues are available from the publisher at the current single issue price.

MICROFILM copies of complete volumes are available from Volume 1 onwards from University Microfilms International Ltd, 30–32 Mortimer Street, London W1N 7RA, England, or 300 North Zeeb Road, Ann Arbor, Michigan 48106, U.S.A.

ADVERTISING. For details contact Bill Drake, telephone (0869) 38477, or write c/o the publishers.

U.S. MAILING. Second class postage paid at New York. Postmaster: Send address corrections to: Analysis: c/o Expediters of the Printed Word Ltd, 515 Madison Avenue, New York, NY10022 (U.S. Mailing Agent).

CONTRIBUTIONS. Articles submitted for publication should be addressed to Mr Christopher Kirwan, Exeter College, Oxford OX1 3DP.

Contributors are asked to note the following requirements. Articles should normally not exceed 3,000 words in length, including references and footnotes. They should be type written in double spacing on one side of the paper only; photographic copies are acceptable. Footnotes should be kept to a minimum and wherever possible avoided altogether. Inset quotations and formulae should be well spaced from the text preceding and following them. The spacing of symbols within formulae should be carefully attended to. Single quotation marks should be used, except for purposes of internal quotation and "scare" quotes. Discussion papers should be sent in as soon as possible after the appearance of the article to which they refer.

Typescripts of rejected articles will be returned only if self-addressed envelopes with correct postage are provided (in the case of overseas contributors, international reply coupons sufficient for airmail).

Proofs of accepted articles will be sent to authors for correction, together with information about free copies. Typescripts of accepted articles will be retained by the Editor on the assumption that authors have kept their own copies.

Apart from correction of printers' errors, only very minor changes in proof will normally be accepted by the Editor, to avoid delay and expense in production.

Typeset by Unicus Graphics Ltd, Horsham. Printed and bound in Great Britain by
Whitstable Litho Ltd, Whitstable

ol. 45 No. 4 (New Series No. 208) October 1985

ANALYSIS

Edited by

CHRISTOPHER KIRWAN

CONTENTS

Report on ANALYSIS problem no. 20	MARTIN HOLLIS
Thoughtcrime	WILLIAM KNORPP
Is the future unreasonable?	A. J. DALE
Thinking about the future	V. H. DUDMAN
'If' and quantification	IVOR ALEXANDER
The complex matter of denoting	PAOLO DAU
'I am here now'	GERALD VISION
Reply to Wright on conditionals and transitivity	E. J. LOWE
A quick Read is a wrong Wright	HILARY PUTNAM
Distribution, superposition and quantum logic	JONATHAN HARRISON
Weir and those 'disproofs' I saw before me	NEIL TENNANT
Noonan, 'best candidate' theories and the ship of Theseus	B. J. GARRETT
Reply to van Brakel	D. K. BUCKNER
Green and dictators revisited	DAVID GORDON
Agnosticism	THOMAS V. MORRIS

BASIL BLACKWELL PUBLISHER · OXFORD

ISSN 0003-2638

Mind & Language

a new quarterly journal from Basil Blackwell

CALL FOR PAPERS

In 1986, Basil Blackwell are launching a new journal, *Mind and Language,* which aims to publish original articles by researchers in all disciplines connected with the study of phenomena of mind and language.

Contributors are invited to submit papers now for the first volume of this new journal.

Format and other information

1. Manuscripts should be submitted to:

   Dr Samuel Guttenplan (Executive Editor)
   Mind and Language
   Birkbeck College
   Philosophy Department
   Malet Street
   London WC1E 7HX
   England

2. These should be in triplicate, double spaced throughout, and typed on one side only of white A4 (8½" X 11") paper. Any footnotes should be clearly numbered and follow the text. In a contribution without footnotes, the source should follow quotations in the text in the form: (Schenker 1979:23) and an alphabetical bibliography of all references should follow the text. Translations, with a record of their source, must be provided for all quotations not in English. Where quotations in the original language are considered essential, the translation should be given in a footnote.

3. It is intended that all papers will be seen by two referees, and that decisions will be forwarded within one month of receipt of a submission. Authors may request that their name is withheld from the referees, in which case it is the responsibility of the author to prepare the manuscript in such a way that after detachment of the title page there remains no indication of authorship.

4. Each contributor will receive ten free copies of the journal in which his or her article appeared. Additional copies may be purchased from Basil Blackwell Ltd. 108 Cowley Road, Oxford OX4 1JF, England.

Basil Blackwell

Theory and Understanding

A Critique of Interpretive Social Science

FINN COLLIN

'This is a first-rate piece of work. Collin's overall conception, and a large number of his arguments, are original, and he is always intelligent and insightful.' David Papineau
Collin characterizes and criticizes various approaches to the social sciences in the light of recent work in the philosophy of mind, language and action, concluding that it is impossible to construct a social science that is both interpretive and genuinely theoretical.
400 pages, **£22.50** (0 631 14256 8)

Invitation to Philosophy

MARTIN HOLLIS

'I read it with pleasure and consider it a good introduction to philosophy.' A.J. Ayer
'Lively, attractive and full of arguments. It offers an invitation to philosophy that many people will want to accept.' Bernard Williams
200 pages, hardback **£14.50** (0 631 14225 8)
paperback **£4.95** (0 631 14226 6)

Thinking about Social Thinking

An Introduction to the Philosophy of the Social Sciences

ANTONY FLEW

Antony Flew insists throughout this provocative book that the social sciences have attracted more than their fair share of misconceptions and 'wishful thinking', and attempts to right these wrongs by giving the reader a clear view of the goals and methods of modern social science.
232 pages, hardback **£16.50** (0 631 14189 8)
paperback **£5.95** (0 631 14191 X)

Understanding Social Science

A Philosophical Introduction to the Social Sciences

ROGER TRIGG

This introductory textbook for anyone studying the nature of social science, questions its scientific basis and shows that philosophical presuppositions are a necessary starting point for any study of society.
272 pages, hardback **£18.50** (0 631 13365 8)
paperback **£5.95** (0 631 14161 8)

The Ivory Tower

ANTHONY KENNY

In these essays Anthony Kenny shows how a philosophical approach can throw light on topics of urgent practical concern. They deal with legal topics, with the relationship between psychiatry and the courts, with the structure of education, and with the issues of nuclear war and nuclear deterrence.
160 pages, **£15.00** (0 631 13985 0)

108 Cowley Road, Oxford OX4 1JF

Suite 1505, 432 Park Avenue South, New York NY 10016

Philosophical Studies

An International Journal for Philosophy in the Analytic Tradition

Editor-in-Chief: JOHN POLLOCK

Associate Editor: KEITH LEHRER

Some Recent Articles
Peter Unger: The Causal Theory of Reference. **Roderick Chisholm:** On the nature of the Psychological. **Alvin Plantinga:** On Existentialism. **David Lewis:** Extrisic Properties. **Charles Chihara:** The semantic Paradoxes: Some Second Thoughts. **Pavel Tichy:** Subjunctive Conditionals: Two parameters vs. Three. **Frank Jackson:** On Indicative Conditionals with Contrary Consequents.

Subscription Information
1985, Volume 47–48 (6 issues)
Institutional rate: Dfl. 185,–/US$ 70.00 per volume incl. postage and handling
Private rate: Dfl. 70,–/US$ 26.50 per volume incl. postage and handling
Private subscriptions should be sent direct to the publishers

D. REIDEL PUBLISHING COMPANY

A Member of the Kluwer Academic Publishers Group

P.O. Box 17, 3300 AA Dordrecht, The Netherlands
190 Old Derby Street, Hingham MA 02043, U.S.A.
Falcon House, Queen Square, Lancaster LA1 1RN, U.K.

Law and Philosophy

An International Journal for Jurisprudence and Legal Philosophy

Edited by ALAN MABE
Dept. of Philosophy, Florida State University, Tallahassee, U.S.A.

Law and Philosophy is a forum for work in law and philosophy that is of common interest to members of the two disciplines of Jurisprudence and Legal Philosophy.

Subscription Information ISSN 0167-5249
1985, Volume 4 (3 issues)
Institutional rate: Dfl. 185,–/US$ 70.00 per volume incl. postage/handling
Private rate: Dfl. 72,–/US$ 27.50 per volume incl. postage/handling
Private subscriptions should be sent direct to the publishers.

D. REIDEL PUBLISHING COMPANY

A Member of the Kluwer Academic Publishers Group

P.O. Box 17, 3300 AA Dordrecht, The Netherlands
190 Old Derby Street, Hingham MA 02043, U.S.A.
Falcon House, Queen Square, Lancaster LA1 1RN, U.K.

METAPHILOSOPHY

Vol. 16 : 1 January 1985

On finding one's feet in philosophy: from Wittgenstein to Marx — KAI NIELSEN

Sentence meaning and speech acts — MICHAEL EDWARDS and JERROLD J. KATZ

The generalized argument from verification: work towards the metaepistemology of perception — SCOTT P. ROBERTS

Is philosophical anthropology possible? — H.P. RICKMAN

Engineering and the philosophy of science — TAFT H. BROOME Jr

Subjectivism and the question of social criticism — ANTHONY WESTON

Notes

Philosophers and experimental inquiry: a reply to Milgram — MARIO MORELLI

1985 Subscription Rates
Institutions: £32.00 (UK) £40.00 (overseas) $75.00 (US) $91.20 (Canada)
Individuals: £18.00 (UK) £22.50 (overseas) $41.80 (US) $51.30 (Canada)

Orders, with payment, to: Iris Taylor, Journals Department, Basil Blackwell, 108 Cowley Road, Oxford OX4 1JF, England.

Teaching Philosophy

Contents October, 1985 Vol. 8/4

James Van Evra	*Logic, the Liberal Science*
John McPeck	*Critical Thinking and Trivial Pursuit*
Daniel Cohen	*Putting Paradoxes to Pedagogical Use in Philosophy*
James Moor & Jack Nelson	*BERTIE II: Personal Computers and Logic*
Richard McCarty	*Media Projects for Elementary Logic*
Stephen Fishman	*Writing-to-Learn in Philosophy*
Sarah Steuber-Bishop	*Telling Tales out of School: Feminist Integration of Curricula*

Subscriptions (4 issues): $16 individuals; $35 others. Add $2 for postage outside the USA. Send orders to the Philosophy Documentation Center, Bowling Green State University, Bowling Green, OH 43403-0189.

Ed by Arnold Wilson, University of Cincinnati, Cincinnati, OH 45221-0206.

ANALYSIS 45.4 OCTOBER 1985

REPORT ON *ANALYSIS* 'PROBLEM' NO. 20

By Martin Hollis

SETTING an Orwellian 'Problem' in 1984 seemed a good idea at the time but, to judge by the number of entries, turned out to have been doubleplusungood. Only four entries were received and only William Knorpp, who wins my applause and £50, scored a substantial point.

The problem was posed thus:

> Syme bit off another fragment of the dark-coloured bread, chewed it briefly, and went on:
>
> Don't you see that the whole aim of Newspeak is to narrow the range of thought? In the end we shall make thoughtcrime literally impossible, because there will be no words in which to express it. Every concept that can ever be needed, will be expressed by exactly *one* word, with its meaning rigidly defined and all its subsidiary meanings rubbed out and forgotten. (*Nineteen Eighty-four*, part I, chapter 5)
>
> The principles of Newspeak, and especially the division of words into three distinct classes, are described by George Orwell in the appendix to the book. Does the project sketched by Syme have any philosophical snag which could not be remedied?

The question is not hard, if it is answer enough that, since whatever can be asserted can be denied, thoughtcrime is possible in any language with negation. As one entrant, Richard Farr of Cornell, noted,

So our powers of distinction might rot
and our vocab. could all go to pot;
but this wouldn't show
that we'd ceased to know
the invaluable function of 'not'.

But this does not touch the relation of thought to language and it gives the arts of rhetoric and propaganda too little credit.

Orwell rested Newspeak on a caricature of a familiar empiricism, which would allow a language with only (A) words for everyday 'impressions and ideas', (B) bellyfeel expressions of attitude and (C) technical terms. Granted also some Whorfian hypothesis that people's experience is shaped by their language, thought control should be possible. Tom Paine claimed optimistically in *The Rights of Man* that 'though man may be kept ignorant, he cannot be *made* ignorant'; but I fear that Dr Goebbels and a cadre of guardians could make a fair attempt. The test would be how thoroughly they could starve experience by putting blocks on the reasoning needed to make sense of a richer heaven and earth and to support subversive moral attitudes. Ideologies can be sustained by comfort as well as by force.

The final solution would lie, I agree with William Knorpp, in the withering away of the state. That is on the cards, only if Oldspeak can shrink into Newspeak in principle. Obvious questions arise about, for instance, the basis and scope of computer languages, the true grammar of ethics, the Whorf hypothesis and the role of language in political control. They lead, I think, to a different theory of meaning from the familiar empirism of *1984*. But an improved philosophy of language might encourage a more potent Newspeak – hence the 'Problem' set.

Although only four entries met the terms of the competition, an interesting missive arrived from Professor Dennis Rohatyn, editor of *Insight, Prophesy and Moral Vision: Critical Essays on George Orwell* (*Cogito*, Vol. I, No. 3/4, September 1983).

THOUGHTCRIME

By WILLIAM KNORPP

ALAS, poor Syme, it seems that he has worked at the Ministry of Truth too long. The idea that Newspeak can make thoughtcrime 'literally impossible' is not only inconsistent with the structure of Newspeak itself, but also contains a conceptual impossibility.

The contradiction between Newspeak structure and Syme's claim that Newspeak can make thoughtcrime impossible becomes apparent when the three Newspeak vocabularies are examined: the 'A' vocabulary for use by the common people; the 'B' vocabulary for members of the Party; and the 'C' vocabulary for scientists. Now, consider the following: if Newspeak theory is to be taken seriously – that is, if we are to believe that the range of thought can actually be narrowed by narrowing the range of language, and that it will then become *literally impossible* for affected individuals to think certain thoughts – then we must ask what function the 'B' vocabulary can possibly serve. None, surely, if we take the theory seriously, for once the 'A' vocabulary has been established, all that would be necessary for the users of the 'B' vocabulary to do in order that *no* thoughtcrime would *ever* again occur is not to pass on their language to any other persons. Is this what Orwell had in mind? It seems, rather, that the 'B' group is intended to exist for as long as the state does (hence, at least in theory, as long as humanity – such as it is – exists). Apparently, however, belief in Newspeak theory was none too strong in its creator, as the group which is supposedly entirely incapable of committing thoughtcrime is envisioned as being policed by another group – surely, then, it is understood that thoughtcrime will not be literally impossible. In fact, if the theory was taken to its logical extreme, the 'B's would be the only group *capable* of committing thoughtcrime, as they are the only ones who possess the concept and, hence, constitute Newspeak's weakest link. It seems that the theory can – or, rather, must – do without either the claim that thoughtcrime will become impossible, or the possessors of the 'B' vocabulary.

But, despite these internal inconsistencies, let us grant to the 1984-state the power to establish its vocabularies, and, for that matter, eliminate thoughtcrime entirely. Now, first we should realize that simply eliminating thoughtcrime is not sufficient for Newspeak to achieve its goal – rather thoughtcrime must be made *impossible*. The conceptual impossibility surfaces when we realize that, in order for the 1984-stage to *be* a state, it must allow some actions and disallow others – simply stated, so long as the state exists, it must exist *as opposed to* the possibility of something else, and so long as this opposing possibility exists (and, simultaneously, defines the state) Newspeak's goal has not been achieved. In actuality, then, the concept of thoughtcrime is *necessary* for the state to exist – without the existence of entities *excluded from* the state, there can be no entities which are *included in* the state in any meaningful way, but only in the trivial sense that *everything* might be assimilated into the concept of state, thus not eliminating the existence of thoughtcrime-like things, but only causing those things to no longer be called thoughtcrimes. The conceptual problem is thus something akin to this: can a meaningful legal system be said to exist where there is nothing which is unlawful? The idea seems absurd.

Southeast Missouri State University © WILLIAM KNORPP 1985

IS THE FUTURE UNREASONABLE?

By A. J. DALE

IN a recent paper V. H. Dudman claims that the division of sentences containing 'if' clauses into indicative and subjunctive fails to conform to a more fundamental distinction between the messages encoded by such sentences ('Parsing "if" sentences', ANALYSIS 44.4, October 1984, pp. 145-53). To illustrate his thesis he gives the following sentences:

(1) If Oswald did not shoot Kennedy, someone else did.
(2) If Hitler invades England, Germany will win the war.
(3) If Hitler invaded England, Germany would win the war.
(4) If Hitler had invaded England, Germany would have won the war.

'Contemporary wisdom' classifies (1) and (2) together as indicative conditionals while (3) and (4) are classified together as subjunctive conditionals (though the choice of the terms 'indicative' and 'subjunctive' may not be entirely happy as Smiley has pointed out). Dudman, on the other hand, wishes to separate (1) from the other three, pointing to a grammatical difference between it and them. In (1), alone of the four, the 'if' clause contains a verb the tense of which matches the notional time implied by the clause: such grammatical differences, he claims, reveal and correspond to differences in the message encoded. (In accord with present conventions attention to grammar could also lead to a distinction between the messages encoded in indicative and subjunctive conditionals, though attention then would centre on the presence of the auxiliary 'would' in the main clause).

Dudman explains the message encoded in sentences where the formal tenses employed match the notional time of their clauses as *hypothetical*: the 'if' clause of such a sentence is understood as a hypothesis:

> ... a speaker who asserts 'if P, Q' under a hypothetical interpretation intimates that he somehow concludes that Q from the hypothesis that P, perhaps because his commitments include some proposition G such that the inference with G and P as joint premises and Q as conclusion is valid. (op. cit. p. 148)
>
> The speaker who ventures (1) is owning to some such argument as 'Someone shot Kennedy; Oswald didn't; therefore someone else did'. (ibid. p. 153)

In contrast, a sentence like (2), (3) and (4) where the formal tense of the 'if' clause does not match its notional time calls for what Dudman terms a conditional interpretation which

> ... invokes the imagined satisfaction of its subordinate condition: to formulate such a judgement the speaker is obliged to conduct a *fantasy* in which the 'if' condition is satisfied. (ibid. p. 150)

He concludes, however, that

> hypotheticals and conditions are ... products of quite different styles of reasoning. A hypothetical is arrived at by arguing from proposition to proposition, a conditional by envisaging a developing sequence of events. (ibid. p. 153)

While willing to admit a fantasy life parallelled only by Walter Mitty's, I am not convinced that the use of 'if' clauses in speech or thought adds to it. In particular I shall argue that many, if not all, 'if' sentences concerning the future conform to Dudman's analysis of the hypothetical interpretation despite their surface grammar.

There are the following two objections that could be made against the hypothetical interpretation as an interpretation of any 'if' sentence.

Firstly, it would support material implication as a *bona fide* representation of such sentences. For the argument $\sim P \vee Q$; P; therefore Q is – *pace* Belnap – valid, and thus a speaker could correctly assert 'if P, Q' solely on the ground $\sim P \vee Q$ (i.e. $P \supset Q$). It may be that Dudman accepts this conclusion for his hypothetical interpretation but there is no indication of it in his paper.

Secondly, it is surely just false to maintain that someone asserting 'if P, Q' *concludes* Q from the hypothesis P, for that would imply the speaker's commitment to Q which fits oddly with his assertion of 'if P, Q' under standard conversational conventions.

Because of the second of the above objections, I do not accept Dudman's representation of the hypothetical interpretation. I shall therefore present an *ad hominem* argument against his claims throughout the remainder of this paper.

The worst consequence of such a view is that it becomes impossible to reason about the future in the same way as about the past. For it would seem to be a grammatical rule of English that 'if' sentences concerning the future must employ the present tense for the 'if' clause. So the formal tense of such a clause can never match its notional time. (Dudman states that this feature is shared by 'when' clauses in such sentences as:

When the alarm sounds I will ignore it

and thus that here too the imagined satisfaction of its subordinate condition is invoked. Presumably, French speakers are restricted in their fantasies since the grammar of French would demand the use of the future tense in such contexts – but then, they order this matter better in France.) Consider the sentence:

(5) If Bush does not win the next presidential election someone else will.

The formal tense of the 'if' clause does not match its notional time so that on Dudman's analysis we do not have an example of a hypothetical. But surely (5) fits as easily as (1) into Dudman's condensed argument account of the hypothetical interpretation: someone will win the next presidential election; it will not be Bush, therefore it will be someone else. This seems exactly on a par with Dudman's view of (1), yet the corresponding 'if' sentence in this case must alter the formal tense. There seems to be no way of 'encoding' the message, to use Dudman's phraseology, correctly. Even in his most convincing example:

(6) If the surgeon does not operate the patient will die,

the speaker may be owning to some such argument as: (to the speaker's knowledge) all in the patient's condition without an operation have died; the surgeon will not operate; therefore the patient will die. With the passage of time and no relevant addition to his knowledge, the speaker may offer:

(7) If the surgeon did not operate then the patient died

and own to the argument: all in the patient's condition without an operation have died; the surgeon did not operate, therefore the patient died. In (7) formal tense does match notional time and thus, Dudman would claim, we now have a hypothetical interpretation of the 'if' sentence. What is puzzling is how the passage of time alone could possibly manage to achieve a change in the form of reasoning involved.

Dudman makes much of the fact that (6) does not obviously contrapose, for

(8) If the patient does not die the surgeon will operate

may well differ in truth value from (6): there is in both (6) and (8) an implied order. (I find this in (7) also; Dudman does not: so much for linguistic intuitions.) Because (6) does not contrapose, he claims that it cannot rest on some argument for arguments are 'meet to be upended'. But the tentative argument: all in the patient's condition without an operation have died; the surgeon will not operate; therefore the patient will die – meets an analogous difficulty on 'upending' for it would then become: all in the patient's condition without an operation have died, the patient will not die, therefore the surgeon will operate. This makes little sense as it stands and demands the substitution of 'will have operated' for 'will operate' to make its sense clear. But then the corresponding 'if' sentence would then be:

(9) if the patient does not die, the surgeon will have operated.

(9) would then be a possible candidate for the expression of the contrapositive of (8). I do not wish to commit myself on the issue of what the contrapositive of (8) is, but it would seem that the difference between the formal tense and the notional time in an 'if' clause concerning the future has little to do with the availability of some supporting argument. Apparent failure of contraposition will be mirrored by an analogous apparent failure in the corresponding argument: it does not show that there is no such argument.

Finally Dudman's own analysis of the sentence:

(10) If this solution turns green when I add the reagent, the deceased died of hyocine poisoning

in terms of the conditional interpretation is at odds with his views on contraposition, for surely (10) is the contrapositive of:

(11) If the deceased did not die of hyocine poisoning then the solution will not turn green when I add the reagent.

In (11) there is a match between the formal tense and the notional time of the 'if' clause and thus it falls under Dudman's hypothetical interpretation. So it should contrapose. Unfortunately the contrapositive is (10), a sentence which falls under the conditional interpretation and thus does not contrapose ...

The University,
Hull HU6 7RX

© A. J. DALE 1985

THINKING ABOUT THE FUTURE

By V. H. DUDMAN

THE natural interpretations of (1) and (2) are products of quite different kinds of thinking about the past. Now, which of these two kinds of thinking underwrites the natural, future interpretation of (3)?

(1) If Oswald did not shoot Kennedy someone else did
(2) If Oswald had not shot Kennedy someone else would have
(3) If Oswald does not shoot Kennedy someone else will.

Philosophers at large tend to think the first; I the second. A. J. Dale ('Is the Future Unreasonable?' ANALYSIS 45.4, October 1985, pp. 179–83) supports the former view, and I am grateful to him for bringing me back to this issue.

It will be observed that (3) differs from (2) only in point of syntactical tense: we have 'will' and '*V*-s' instead of 'would' + phase and '*V*-ed' + phase. We should therefore expect the informational differences between their encoded messages to reduce to a purely temporal one. I account it among the consolations of my theory that this is exactly how matters pan out. For my theory is that the messages of (2) and (3) are both arrived at by imagining futurewards from certain historical realities of a temporal point *v*, and that the only difference between them is that (2) registers *v* as past with respect to some already past point, whereas (3) identifies *v* as the point of speech. In my submission, then, the messages of (2) and (3) have a common logical form.

This logical form is established as follows. Each of the natural, future interpretations of (4) and (5) is a different *judgment* about the same topic, Kennedy's being shot by 'someone else':

(4) Someone else will shoot Kennedy
(5) Someone else can/may/ought to/ must shoot Kennedy.

Patently, the judgment itself is registered by the secondary auxiliary. So we can crudely depict the natural interpretation of (4) as

will (*someone else, shoot Kennedy*)

with boldface 'will' representing whatever informational factor it is that conditions the selection of lexical WILL in preference e.g. to MUST. Now, when (4) and (5) are each expanded to incorporate the subordinate clause 'if Oswald does not shoot Kennedy', what common complication of their original messages does this 'if'-clause encode? The answer is that the *judgment itself* becomes modified: in the message of (3), the operator of widest scope is registered by 'will if Oswald does not shoot Kennedy':

(C) [**will** *if not* (*Oswald, shoot Kennedy*)] (*someone else, shoot Kennedy*).

The messages of (5), (4) and (3) are all equally judgments about someone other than Oswald shooting Kennedy in future. And I claim that each of these different judgments is arrived at after contemplating a future situation imagined to eventuate as a consequence of certain historical realities of the point of speech, adding that in (3)'s case the imagined development of this future situation is a sequence of events from which occasions of Oswald's shooting Kennedy are definitely excluded.

And so to the business of tense. I propose that we arrive at the past interpretation of (2) exactly as we arrive at the future interpretation of (3), except for starting from a past past *v*. Accordingly (C), above, represents their common logical form.

And the thinking common to them both comprises two phases: a phase in which subsequent developments are spun imaginatively out of the prior historical realities, and a phase in which the speaker, contemplating those imagined developments, opts for one or another judgement concerning the topic of someone else's shooting Kennedy, thereby conditioning the choice between the modals WILL, MUST, CAN, etc.

While (3) is generated by insertion of an entire 'if'-clause into the predicate of (4), (1) is rather generated by juxtaposing the separate sentences 'Oswald did not shoot Kennedy' and 'someone else shot Kennedy'. The propositions that Oswald did not shoot Kennedy and that someone other than Oswald shot Kennedy are separate semantic components of the natural interpretation of (1), which I analyse as a *condensed argument* with the proposition that someone shot Kennedy as its ground:

(6) Someone shot Kennedy, Oswald did not shoot Kennedy; therefore someone else shot Kennedy.

Behind the natural interpretation of (1) lies a piece of deductive reasoning.

In sum, my position is that the messages of (2) and (3) are both *judgments* about *imagined* situations, and quite unlike the message of (1), which is a condensed *demonstrative argument.*

Dale reckons it a consequence of this position that ‘it becomes impossible to reason about the future in the same way as about the past’ (op. cit. p. 181); but this is a faulty appreciation, for he says nothing to impugn my doctrine that we arrive at the future interpretation of (3) exactly as we arrive at the past interpretation of (2). His complaint turns out upon investigation to be that my doctrine repudiates a similarity he thinks to descry between thinking leading to a broadcast of (3) and thinking leading to a broadcast of (1). For Dale sees the message of (3) as underwritten by an argument, just like (1)’s only relating to the future (op. cit. p. 181):

(7) Someone will shoot Kennedy; Oswald will not shoot Kennedy; therefore someone else will shoot Kennedy.

Actually, Dale’s own example at this point concerns the winning of the next presidential election. *If any*, of course.

Now I allow that a speaker who already independently concedes that Kennedy will be shot tomorrow is entitled to conclude that if Oswald does not shoot Kennedy tomorrow someone else will. But this, I maintain, is not because the thinking behind (3) is the demonstrative reasoning unfurled in (7), but merely because the imagined development's outcome has been prejudged.

I also allow that, in virtue of the demonstrative cogency of the reasoning unfurled in (7), a speaker who for whatever reason independently concedes that Kennedy will be shot tomorrow can venture (8) as a condensed argument:

(8) If Oswald (being bound to miss, e.g., or already locked up) will not shoot Kennedy tomorrow, then someone else will shoot Kennedy tomorrow.

But then, (8) is not (3).

And really, this is the end of the line for Dale. For on the one hand, the messages of (3) and (8) are recognisably *different*, and on the other, the condensed-argument interpretation underwritten by the reasoning in (7) is recognisably (8)’s. Therefore it is not (3)’s; q.e.d.

Those for whom the message of (3) *is* underwritten by the reasoning in (7) are of course obliged to analyse ‘Oswald does not shoot Kennedy’ in (3) as encoding the natural, future interpretation of

(9) Oswald will not shoot Kennedy.

And so generally: upon those who conceive future conditionals as condensed arguments, the doctrine is incumbent that English puts ‘*V*-s’ *instead of* ‘will’ in future-referring ‘if’-clauses. Willy nilly, then, they find themselves committed to *grammatical* dogma. Archetypal philosopher, Dale embraces the incumbent grammatical tenet, advancing in its favour the spurious inducement that the speaker of English is 'obliged', for some unexplained reason, to resort to the unpredictable locution:

> it would seem to be a grammatical rule of English that 'if' sentences concerning the future must employ the present tense for the 'if' clause (op. cit. p. 181).

Now, 'it would seem to be' would seem to be a strangely guarded way of heralding what after all is a central plank of the entire philosophical platform. But anyway, the doctrine is untenable, exploded in the previous paragraph. The natural interpretation of (3) cannot be a condensed argument of any kind, because there is nothing encoded into (3) which might serve as a premiss. That premiss cannot, witness (8), be the natural interpretation of (9); yet no other premiss will do. And so generally: by whatever thinking we arrive at future conditionals, it cannot be by drawing conclusions from premisses.

My response to Dale, then, is that his rival analysis of future conditionals is overthrown by considerations, purely empirical, of English grammar.[1] And this, to my mind, exactly epitomises the relation of logic to grammar. For when all is said and done, a theory of predication is per se a *grammatical* theory.

Macquarie University,
North Ryde, N.S.W.
Australia, 2113.

© V. H. DUDMAN 1985

---

[1] The framing of this response owes much to the guidance of D. H. Mellor.

'IF' AND QUANTIFICATION

By IVOR ALEXANDER

V. H. DUDMAN contends in his article 'Parsing "If"-Sentences' (ANALYSIS 44.4, October 1984, pp. 145–53) that the sentence

(1) If Baby cries we beat him

is not equivalent to

(2) $\forall t$ (if Baby cries at t, we beat him after t).

He writes: '[(2)] simply makes no sense, for "t" would have to range over elements of McTaggart's *A*-series, yet "Baby cries ten minutes ago" is not an English sentence' (p. 149).

The argument seems to be as follows: (1) is quite meaningful, so if (2) is equivalent to (1) then (2) is meaningful as well. If (2) is meaningful then so is every instantiation from

(3) Baby cries at *t*.

The variable '*t*' must range over the *A*-series. Therefore

(4) Baby cries ten minutes ago

is one such instantiation. But (4) is not meaningful, etc.

Will the argument hold? We will endeavour to show three things. First, (4) is not in fact an instantiation from (3). Second, '*t*' ranges over elements of the *B*-series. And third, (2) is not weaker than (1).

I

Dudman is of course right when he says that (4) is not an English sentence, but he has neglected to explain why. A relevant answer might be offered as follows.

A. A verb occurrence that comprises essentially an infinitival stem and an '-s' suffix may be untensed. Take for example 'has' in the sentence 'Seventeen has no integral divisors other than ±1, ±17'. (The thesis is not novel, and Dudman has not rejected it, but we state it explicitly.) Or consider the modern history teacher who asserts

(5) Columbus discovers America in 1492.

To appreciate that the verb here is untensed we have only to acknowledge either of two things. First, if in fact he were speaking in the present tense then the teacher would be trying to obscure the pastness of the event in question (relative to everyone in the classroom). But this is not his aim. And second, no verb occurrence in the true present tense is qualified by a temporal adverb. An allusion to the moment of utterance is sufficient. (In this latter regard suppose the same teacher looks out the window, sees his pupil Higgins approaching across the fields, and says 'Higgins is arriving late'. The presentness of the action reported is palpable. But if he adds (e.g.) 'on the first day of term' then the verb that occurs in the present tense is no longer *to arrive late*, and has become *to arrive late on the first day of term*. Some students have never done it, but Higgins is doing it now. Only a person living in 1492 could speak analogously about Columbus.)

B. Truly tensed verb occurrences are egocentric (or 'noscentric'). If in another context the teacher says

(6) Columbus discovered America in 1492,

then he communicates to his listeners not only (i) that the date of Columbus's accomplishment is [tenselessly] 1492, which is all that may be inferred from (5), but also (ii) that the discovery occurs earlier in the temporal continuum than his relating of it. The later time alluded to is the Now of the classroom, and the binary relation obtaining between the two events is signalled quite expressly by the '-ed' suffix on the verb. There is in fact no more suggestion of pastness in (5) *per se* than there is in

(7) The Americans vote in 1992.

The reasons for the ungrammatical status of (4) are now plain enough. To begin with, 'cries' as it occurs therein is either tensed or untensed. If it is tensed, then, since it is also indicative, it is simple present or present habitual. The second possibility may be discounted, since (1) is not here construed as equivalent to 'If Baby makes a habit of crying we beat him'. If it is simple present it is inconsistent with the pastness implicit in its qualifier, i.e. the adverbial phrase 'ten minutes ago'. If on the other hand it is untensed, and hence objective, then it is inconsistent not only with the pastness of that same adverb but also with its egocentricity.

The importance for us of this ungrammaticalness is that (i) it is rectifiable and (ii) it does not derive from some fault intrinsic to (2). We take the two points separately.

First, we can fix (4) in either of two ways. The choice will depend on how much information the speaker wishes to impart. If he wishes to leave himself (or his audience) temporally out of the matter then he will say (e.g.)

(8) Baby cries at 11:04 pm on 9 May 1987,

where the precise time in question may be earlier than, identical with, or later than the moment of speech. But if his inclination is to point out not only (i) the time of the incident but also (ii) his temporal relation with it, then he will say

(9) Baby is going to cry at 11:04 pm on 9 May 1987,

or

(10) Baby is crying,

or

(11) Baby cried ten minutes ago,

and so on. The adverb occurring in (9) is objective, in the case of (10) there is none necessary (since the time of the utterance itself is alluded to), and in (11) it is egocentric. But in each example the verb is tensed, hence the sentence in which it occurs is predominantly subjective.

Second, if (2) is to be interpreted as an explication of (1) then consistency requires that whatever constraints apply to (1) shall apply also to (2). One such constraint is imposed by English grammar, according to which (4) does not describe one of the conditions generalized in (1). Hence there is no reason why we should deem it to be an instantiation from (3). Such instantiations are invariably like (8), i.e. untensed and objectively dated. With '*T*' as an arbitrary objective date we may say that in general they are of the form

(12) Baby cries at *T*,

to which no more mystery attaches than to (5) or (7).

II

McTaggart defines the *A*-series as 'that series of positions which runs from the far past through the near past to the present, and then from the present through the near future to the far future, or conversely' (J. M. E. McTaggart, *The Nature of Existence*, Volume II, Cambridge University Press, 1927, Book V, Chapter 33, p. 10). Since the *A*-series is therefore subjective, and (12) is not, '*t*' as it occurs in (2) cannot be understood to range over that same series.

But consider the *B*-series. It is defined as running from earlier to later or conversely (ibid.), so that it is ordered by a transitive asymmetric relation and will map onto the real number line. A further attraction is its objectivity, by virtue of which it is compatible with (2). In order to see that it is, in fact, the domain of '*t*', we have only to remember that in every sentence like (12) the time of the reported incident is expressed objectively: any given '*T*' is a position in the *B*-series.

We now consider two possible objections to the thesis that (2) is not weaker than (1). First, it might be objected that 'we' as it occurs in (2) entails a subjective view of time. 'We' is egocentric and therefore has an affinity with such adverbs as 'ten minutes ago', which plainly call for the *A*-series. Furthermore, each of us may at any moment perceive himself as wedged in between two vast expanses of the continuum, the Past and the Future. Positions in either direction are ordered by their remoteness, which is relative to the same subject. The Present between the two regions is intensely subjective. In the light of this insistent subjectivity it is natural to read '*t*' as ranging over the *A*-series.

The objection fails to acknowledge that the polarity at issue (subjective/objective) relates only to verbs and adverbs. A sentence in which 'we' (or 'I') occurs essentially may be just as tenseless as (5), (7), or (8). (And a sentence in which 'we' does not occur at all may obviously be tensed, as are (6), (9), etc.) To this extent the occurrence of 'we' in (2) has no bearing on the matter.

Second, an objection might be suggested as follows. From (1) and the additional premiss

(13) Baby will cry next week,

for example, we can infer

(14) Baby will be beaten next week or later.

But the *B*-series knows nothing about 'next week', just as it knows nothing about 'ten minutes ago'. Accordingly (2) is not strong enough to justify the inference from (13) to (14), and for inferential purposes it is weaker than (1). Dudman's repudiation of (2) is thus vindicated.

The argument will not work. Equivalent to (2) is the sentence

(15) $\forall t$ (if Baby cries at t then there exists u such that (i) t is earlier than u and (ii) we beat Baby at u).

Furthermore, (13) is equivalent to

(16) There exists t' such that (i) t' is part of next week and (ii) Baby cries at t'.

Now from (15) we know that if Baby cries at 't'' then he is beaten at some subsequent moment, say 't'''. So within the scope of (16) we can infer

(17) There exists t'' such that (i) t' is earlier than t'' and (ii) we beat Baby at t''.

In making this inference we have not availed ourselves of the information (carried by (16)) that 't'' is part of next week. With that and (17) we have all we need to produce (14).

A generalization of this latter argument will establish that (2) is not weaker than (1).[1]

157 Dublin Road,
Sutton, Dublin 13, Ireland

© IVOR ALEXANDER 1985

---

[1] I am grateful to the Editor for helpful advice on this note.

THE COMPLEX MATTER OF DENOTING

By PAOLO DAU

IN 'On Denoting' (*OD*) Russell reports finding 'certain rather curious difficulties' with theories that assign denoting phrases both a meaning and a denotation. The passages detailing these difficulties ([12], pp. 111–13) have generated considerable attention and discussion over the past three decades. By now it is widely acknowledged that little progress can be made in the interpretation of these passages without extensive use of Russell's earlier theory of denoting as set out in *The Principles of Mathematics* (*PoM*: [7]).

But it is not sufficiently well understood, in my opinion, that progress also requires familiarity with Russell's unpublished manuscripts from the period 1903–1905. In my view even when it is supplemented by the doctrines of *PoM* the text of *OD* is simply *inconclusive* with respect to several important interpretative questions. While not all such questions can be settled by reference to the unpublished manuscripts, I believe that some can be. A recent paper in ANALYSIS by Dilip K. Basu ([1]) provides the occasion for a much needed start in this direction. In what follows I shall consider some of the claims he makes, set out my reasons for rejecting them and cite some of the textual evidence in favour of my position. To keep the discussion manageably brief, I have focused on just one issue, to which the relevant texts can be applied in a straightforward and direct manner, namely the interpretation of Russell's 'denoting complex'.

I. DENOTING COMPLEXES

The expression 'denoting complex' is much in evidence in the critical portion of *OD* mentioned above. In a note published in 1959, P. T. Geach claimed that the term is 'a technicality of the [*PoM*] theory' ([5], p. 69). Geach is mistaken about the origins of the term. *PoM* contains no use of the expression (although it does employ 'denoting concept'). At the other extreme, Basu maintains that 'denoting complex' 'occurs in *OD* alone and only in those parts where Russell has Frege under fire' ([1], p. 65). Basu thinks that the term was coined by Russell expressly for the purposes of his critical discussion (ibid. pp. 67–8). This is certainly false. Examination of the manuscripts dating from 1904 and 1905 reveals protracted and detailed discussion of denoting complexes. Far from being coined expressly for the occasion, the term 'denoting complex' was used extensively by Russell in elaborating his own account.

Basu provides a list of the rival interpretations of Russell's denoting complexes, which we may summarize as follows:

(i) A denoting complex is a denoting concept, in the *PoM* sense.
(ii) A denoting complex is some version or other of a Fregean sense.
(iii) A denoting complex is something that denotes a meaning: it is itself neither a meaning, nor a denoting phrase, nor a denotation.

Basu rejects all of these in favour of his own suggestion:

(iv) A denoting complex 'is an expression with a putative meaning and a denotation' (ibid. p. 67).

None of these suggestions is quite correct. I will take them up in the order given.

II. The *PoM* Interpretation

It is important to keep in mind that Russell's first account of denoting underwent subtle but important changes before it was abandoned in *OD* in favour of the new theory. In particular we must guard against the assumption that Russell's views about denoting were fixed, precise and well defined in early 1903, when *PoM* was published. In fact the theory commonly ascribed to Russell, with its apparatus of denoting concepts and denoted objects, was quite a late addition to the text. For example denoting concepts do not even appear in the first versions of the theory, and were only introduced in the final draft of May, 1902; the entire discussion of the definite description operator was also only formulated in the final draft; and some of the examples Russell uses in his discussion of denoting in Chapter V of *PoM* were actually added in proof. (A detailed discussion of the matter can be found in Blackwell [3] and Dau [4].) My point here is just that Russell's thought on the matter of denoting was very much in a state of flux in 1903. It is important to keep this in mind in assessing the first of the four interpretations cited above, which is, in my opinion, the closest to the correct account.

The first interpretation identifies *PoM*'s denoting concepts and denoting complexes. It will be easier to see why it is wrong to make this identification if we begin with what may seem a trivial point: a denoting complex is, first, something that denotes; and second, something that is complex. Putting matters this way suggests that there should be not only *non-denoting complexes* but also what we may call *denoting simplexes.* The suggestion correctly describes Russell's position at different points in the development of the theory of denoting. Indeed the existence of denoting simplexes can even be established by a careful reading of *PoM*, since they play an important (albeit somewhat obscure) role in that work. According to *PoM*, common noun phrases indicate class-concepts. Associated with an expression of this sort is a series of phrases formed by the application of one of the determiners 'all', 'every', 'any', 'a', 'some' and 'the'. These phrases also indicate concepts, and, as is well known, according to Russell the concepts in question all have the property of denoting. But they are not the only denoting concepts in the *PoM* theory. At various places in the text Russell observes – quite correctly, by application of the *aboutness* criterion for denoting which he gives us (*PoM* p. 53) – that class-concepts denote (e.g., *PoM* pp. 54, 55, 152, *inter alia*); furthermore, the indication of plural forms of common noun phrases – what Russell calls concepts of a class – *also* denote (*PoM* pp. 67, 80, 106, *inter* *alia*). In short in *PoM* not just phrases but individual words of a certain kind indicate denoting concepts, and among these concepts some are 'philosophically simple', according to Russell.

(An argument for the existence of denoting simplexes might also be made by appealing to what Russell has to say in *PoM* about the unrestricted variable. On this topic, however, the text is considerably more obscure. In the unpublished manuscript 'On Meaning and Denotation' Russell states his position more clearly. Employing the quotation convention he introduces in *OD* to form meaning-names (*OD* p. 111) he tells us explicitly that a variable, say 'x', 'means "anything" and denotes anything, i.e. denotes what "anything" denotes' ([10], p. 36).)

At the other extreme it is clear that Russell earlier considered to be denoting certain complexes which, by late 1904, he no longer counted as such. A perfect illustration is provided by Russell's analysis of sentences. According to Russell a (true) sentence has both a meaning (the proposition it indicates) and a denotation (a certain fact). 'On Meaning and Denotation' contains the following characterization of Russell's position:

> A phrase which asserts something seems both to express a meaning and to designate what we may call a *fact*. At any rate something of this kind is the case, often if not always, with phrases which assert *true* propositions; with such as assert false ones there is reason to think that the designation has disappeared. If this be the case with propositions, it tends to show that denoting is part of the nature of propositions. ([10], p. 10)

A second unpublished manuscript, 'On the Meaning and Denotation of Phrases', provides a useful illustration of Russell's theory:

> Consider, e.g., 'the present Prime Minister of England is the nephew of the previous Prime Minister of England'. This phrase *denotes* a certain fact, the same fact as is denoted by 'Mr. Arthur Balfour is Lord Salisbury's nephew'. But the *meaning* of the two phrases is quite different. ([9], p. 2)

In short, at this point in his development of the theory of denoting Russell considered true propositions to be complex denoting concepts. It is possible to mount a very strong argument for the thesis that this account of true propositions *follows from* the general *PoM* theory of denoting, although it is not entirely clear whether Russell *recognized* this before mid 1903. (In light of the fact that the theory of denoting of *PoM* was not formulated until mid 1902 it would not be surprising if Russell had not fully recognized some consequences of his position.) What is certain is that, for a variety of good and bad reasons, by late 1904 Russell had changed his mind. 'Propositions', he writes, 'are complexes which do not denote' ([10], p. 63). This remained his view right up to the time *OD* was written. In the June 1905 manuscript, 'On Fundamentals', in which the old theory of denoting is finally abandoned and the new theory proposed in *OD* is first formulated, Russell frequently speaks of propositions as 'undenoting complexes' (e.g. [11], pp. 32, 39).

We are now in the position to state why the identification of denoting complexes and *PoM* denoting concepts is ill advised. The proposal is correct to the extent that any denoting complex is a *PoM* denoting concept, but only to that extent, since the converse claim is false on two counts:

1. Some *PoM* denoting concepts are not complexes.
2. Some *PoM* denoting concepts are later held by Russell to be undenoting complexes.

In light of this, the first interpretation must be reformulated as follows:

(i′) A denoting complex is a denoting concept, in the *PoM* sense, that is complex and non-propositional.

III. THE REMAINING ALTERNATIVES

The second interpretation understands denoting complexes to be some version or other of Fregean senses. Naturally the most interesting form of this proposal is that we *identify* the denoting complex expressed by a phrase and the corresponding Fregean sense. It is worth noting that this further identification is left open by the reformulation of the first interpretation given above. Interpreters who have taken Russell to be mounting a successful criticism of Frege in the passages of *OD* with which we are concerned are committed to this form of the proposal, and are consequently committed to the view that Russell understood Frege's account of sense.

Now Russell himself claims, in *OD*, that his *PoM* theory of denoting 'is very nearly the same as Frege's' (*OD* p. 104 n. 1); but it is nonetheless clear that there are important differences between the two. One of these differences is recognized by Russell. In his published works Frege claims that any grammatically correct proper name has a sense, even if it lacks a reference (*Bedeutung*). According to Russell, on the other hand, no true proper name has both denotation *and* something else. A second difference between the two, more subtle and almost certainly never recognized by Russell, concerns the principles of compositionality that he and Frege each espoused. For Frege the sense of a complex expression depends on – more precisely, is a *function of* – the senses of the constituent parts of the expression; similarly the reference of a complex expression is a function of the references of its parts. Russell too adopts principles of compositionality, but his principles are in an important respect different from Frege's. According to Russell the meaning of a complex expression is literally *composed* of the meanings of its parts: it literally *contains them as constituents.* Similarly, the *denotation* of a complex (denoting) expression is not merely a function of the denotations of its parts but a complex or composite of them. For example, during the time when Russell held that true sentences have both a meaning and a denotation, he maintained that the denotation of the sentence literally contains the denotations of its parts. The following example is typical. Suppose that the phrase 'Smith's wife' denotes Triphena. Then of the sentence 'Smith's wife has blue eyes' we should say, according to Russell,

> that *Smith* and *wife* and the meaning (not the denotation) of 'Smith's wife' are constituents of the total meaning, but none of them are constituents of the denotation, whereas Triphena herself is a constituent of the denotation. ([8], p. 3)

It is obvious that no such account will work for Frege, according to whom the reference/denotation of a sentence is a truth value and not a fact.

However plausible Russell's compositionality principle may be for the denotation of sentences, it is clear that it will not work for descriptions. 'The present Prime Minister of England' does not denote something containing England; and 'the centre of mass of the Solar System at the beginning of the twentieth century' denotes a mathematical point, hence certainly not a complex object containing the entire Solar System or the twentieth century. Of course Russell was perfectly well aware of all this: the examples are his (*cf.* [10], pp. 13, 19). Faced with this difficulty, he was forced to complicate his compositionality principle:

> When a word or phrase which both means and denotes occurs in a phrase, its *meaning* is a constituent of the *meaning* of the whole phrase, and its denotation is a constituent of the *denotation* of the whole phrase or of the meaning, according to circumstances. ([9], p. 1)

In short, Russell was led to the view that the denoting complex expressed by a descriptive phrase typically contains the *denotations* of the parts of the phrase as well as their meanings. For example a denoting phrase of the form 'the centre of mass of x at the time t', Russell writes, 'denotes a point in space, which is a perfectly simple individual; but x and t are both constituents of the meaning which denotes this individual...' ([9], p. 9).

This accounts for Russell's claim, in *OD*, that according to Frege's theory 'the centre of mass of the Solar System at the beginning of the twentieth century'

> is highly complex in *meaning*, but its *denotation* is a certain point, which is simple. The Solar System, the twentieth century, etc., are constituents of the *meaning*; but the denotation has no constituents at all. (*OD* p. 108)

The unpublished manuscripts make perfectly clear that this passage is no mere *slip* by Russell, but instead reveals a pattern of systematic misinterpretation of Frege's theory on his part. Russell's denoting complexes are not Fregean senses. This means that the second interpretation must be rejected. For this reason among others those interpretations which, like Blackburn and Code's [2] and

Hochberg's [6], recommend that we take Russell's criticisms as successfully directed against Frege, are highly suspect. In my view the difficulty Russell means to address in *OD* is actually the result of two theses he did *not* share with Frege, namely his thesis that proper names mean what they denote, and his 'heap' version of the compositionality principle.

After all this, the third interpretation can be disposed of easily. In 1904 Russell discovered a 'curious difficulty' with his *PoM* theory of denoting. In his system a denoting complex can only be spoken about if it is denoted, and not if it is named. This means that, in order to speak about a denoting complex, we must use a phrase expressing a denoting complex with the first complex as its denotation. From this it does not follow, however, that denoting complexes are not themselves meanings, or that they denote *only* meanings, as the third interpretation maintains. The unpublished manuscripts make this completely clear. The following passage is just one of many in which Russell speaks of denoting complexes whose denotations are not meanings:

> ... the combinations of words 'the present Prime Minister of England' and 'the late Lord Salisbury' both express denoting complexes; hence the proposition ['the present Prime Minister of England is a nephew of the late Lord Salisbury'] is also about the two men whom these complexes denote. ([10], p. 25)

Basu's own interpretation is also incorrect. As is well known the interpretation of Russell's early philosophy is made considerably more complicated by his tendency to use key expressions ambiguously. This tendency is also prominent in the unpublished manuscripts, which were, after all, in the main either private notes or drafts meant to be read and discussed only with Whitehead. Like 'proposition' and 'propositional function' in *PoM*, the term 'denoting complex' is also open to alternative readings. Sometimes it is clearly used to refer to notational forms, frequently it is just as clearly used to refer to what these notational forms indicate or express, and sometimes the text does not decide between the two. In spite of this, Basu's proposal that we understand a denoting complex to be a (proper) denoting phrase must be rejected.

The term 'denoting complex' seems to have entered Russell's writings as a short form of 'complex denoting concept' (or 'complex denoting meaning'), and may have been partly inspired by Whitehead's discussion of complexes. In the 1904 work 'On Meaning and Denotation' ([10]), in which the expression makes its first extensive and systematic appearance, it is usually employed to refer to meanings and not to expressions. The theory elaborated in this paper is finally abandoned in 'On Fundamentals' ([11]). 'On Fundamentals' contains some of the critical arguments later repeated in *OD* pp. 111-13, as well as an outline of the new theory set forth in *OD*, 'the theory of denotation *without* meaning' ([11], p. 61).

A striking feature of this manuscript is that as Russell shifts to the 'theory of denotation without meaning', he stops speaking of denoting *complexes*, and returns instead to talk of denoting *phrases.* (Of course the same linguistic turn is evident in *OD* itself.) For these reasons among others it is on balance necessary to take Russell to mean non-linguistic entities by 'denoting complexes'. This interpretation also has the advantage that it avoids the implausible claim, to which Basu is committed, that Russell wishes to maintain that there is some difficulty in speaking of denoting phrases. Incidentally, Basu is also mistaken to suggest, as he does, that Russell was enabled to dispose of denoting concepts by his acquisition – presumably some time in 1903 – of 'Frege's quantification device' ([1], p. 67). Russell already had a version of quantification theory at his disposal in 1900 – he got it from Peano – and indeed *PoM* itself contains a number of passages in which the quantificational paraphrases for (indefinite) denoting phrases later embraced in *OD* are recognized (e.g., *PoM* pp. 36, 89-90, 264-5).

I imagine that it is clear from the above that a great deal more needs to be said about these questions. I hope that it is also clear that a great deal more *can* be said. Russell's unpublished manuscripts offer a rich and important source of information about his thought during the crucial period when he and Whitehead struggled with the Contradiction. In particular, they have a great deal to teach us about the development of the theory of denoting offered in *OD.* At the minimum, as I have tried to show, they can help us establish what the theory is *not* like.

University of California, San Diego,
La Jolla, California 92093,
U.S.A.

© Paolo Dau 1985

References

[1] Dilip K. Basu, 'Russell on Denoting', *Analysis* 43.2 (March 1983), 65-70.

[2] Simon Blackburn and Alan Code, 'The Power of Russell's Criticism of Frege', *Analysis* 38.2 (March 1978), 65-77.

[3] K. Blackwell, 'The Genesis of Russell's First Theory of Denoting', unpublished, 1971.

[4] Paolo Dau, 'Russell's First Theory of Denoting and Quantification', *Notre Dame Journal of Formal Logic*, forthcoming.

[5] P. T. Geach, 'Russell on Meaning and Denoting', *Analysis* 19.3 (January 1959), 69-72.

[6] Herbert Hochberg, 'Russell's Attack on Frege's Theory of Meaning', *Philosophica* (1976), 9-34.

[7] Bertrand Russell, *The Principles of Mathematics*, George Allen & Unwin, 1903, 2nd. edition 1937.

[8] Bertrand Russell, 'Points About Denoting', unpublished manuscript in the Russell Archives (McMaster), 1904.

[9] Betrand Russell, 'On the Meaning and Denotation of Phrases', unpublished manuscript in the Russell Archives (McMaster), 1904.

[10] Bertrand Russell, 'On Meaning and Denotation', unpublished manuscript in the Russell Archives (McMaster), 1904.

[11] Bertrand Russell, 'On Fundamentals', unpublished manuscript in the Russell Archives (McMaster), 1905.

[12] Bertrand Russell, 'On Denoting', *Mind* (1905), reprinted in *Essays in Analysis*, D. Lackey, ed., George Braziller, 1973, pp. 103–119.

'I AM HERE NOW'

By GERALD VISION

> In virtue of its form ['I am here'] must be true on any occasion on which [it is] asserted, and yet the proposition it expresses on each occasion [is] contingent. (Richmond H. Thomason, 'Necessity, Quotation, and Truth: an Indexical Theory', *Language in Focus: Foundations, Methods and Systems*, ed. by Asa Kasher, D. Reidel Publishing Co., 1976, p. 121)

> Intuitively, ['I am here now'] is deeply, and in some sense universally, true. One need only understand the meaning of [it] to know that it cannot be uttered falsely. (David Kaplan, 'On the Logic of Demonstratives', *Contemporary Perspectives in the Philosophy of Language*, ed. by Peter French, Theodore Uehling Jr. & Howard Wettstein, University of Minnesota Press, 1979, p. 402)

> The sentence 'I am here' has the peculiar property that whenever I utter it, it is bound to be true. Even if I am lost and do not know where I am, I can bravely say 'I am here', and know that I am expressing a truth. (Simon Blackburn, *Spreading the Word*, Oxford University Press, 1984, p. 334)

COMMON to these passages is the claim that (it must be the case that) the sentence 'I am here now' is true whenever asserted. Although Thomason and Blackburn omit mention of 'now', it is implicit in their claims; for an indefinite indication of present tense is crucial to whatever plausibility the view may have. Let us entitle the view 'the indexical thesis'. The thesis implies that the sentence even frustrates attempts to lie with it. And as frequently as the thesis appears in print, it is even more often expressed in informal conversation, where I have seldom seen it challenged. Despite this, the thesis is false.

To illustrate, suppose that I am lost in an unfamiliar part of Chicago. I take out my map, put my finger on a certain location and begin saying to myself (or to my companion) 'Let's see, *I am here now* . . .'. In fact, I have pointed to a part of the map corresponding to a mile north of my present location. (What my finger points to is not the same as what my term refers to. My finger points to a green quadrate on a piece of paper. My term refers, say, to a public park.)

I find it difficult to imagine a clearer refutation of this surprisingly popular thesis. Moreover, the counterexample serves against a commonly accepted companion rule for determining in general the reference of 'here': namely, that 'here' denotes the, vaguely bounded, present location of the utterer.

Another basic difficulty for the thesis focuses on an assumption it makes about 'now' (or about indefinitely indicated present tense). That is, the thesis requires the assumption that the time of a statement *must be* its time of utterance. But we commonly send cards reading 'Happy Birthday' or 'Merry Christmas', expecting them to be received at a later, more appropriate, time. Or we leave messages on an answering machine, or notes for acquaintances, informing them of our present whereabouts (viz. our whereabouts at the time the messages are expected to be received, not when they are recorded). If we allow time of utterance and statement to vary for such everyday cases, the following instance should provide an utterance of 'I am here now' that is false.

I come home unexpectedly at midday and go upstairs to rest. My wife is out, but when she comes home at the usual time (say, two hours hence) I suspect that she will sit downstairs waiting for my arrival. To prevent her wasting this time I seek to inform her of my presence by writing a note that begins 'I am here now ...' Of course, I intend it to be read two hours after it is written. But I am awoken after fifteen minutes by a sonic boom, and distractedly leave the premises without picking up my note, which my wife then reads after I have left. If 'now' refers to the time she reads it, consistently with my expectations, what I wrote is false.

In the face of this one might simply insist on the intuitive plausibility of saying that 'now' always refers to the time of utterance (or writing). But if we do not presuppose the very indexical thesis being tested, it is doubtful that this intuition can survive. On the other hand, if a critic is bothered by the similarity between *what* I put on paper two hours earlier and *what* my wife must understand later, imagine that I now leave an audible message on a message machine that contains standard buttons for certain commonly uttered sentences. What I do when I come home is not write a sentence, but simply set the machine to activate two hours hence and press button number ten. Why should this procedure, in which I do not utter or write the sentence token, make it more plausible that 'now' refers to two hours hence than if I had written the sentence on paper?

Temple University,
Philadelphia, Pennsylvania 19122, U.S.A.

© GERALD VISION 1985

REPLY TO WRIGHT ON CONDITIONALS AND TRANSITIVITY

By E. J. LOWE

IN his rejoinder to my earlier article 'Wright *versus* Lewis on the transitivity of counterfactuals' (ANALYSIS 44.4, October 1984, pp. 180–3), Crispin Wright introduces the following new example (see his 'Comment on Lowe', ANALYSIS 44.4, October 1984, pp. 183–5):

> Suppose that I, whom you regard as an expert on the Soviet political system and the kind of characters who succeed within it, inform you that 'If Benn had been born and brought up in Russia, he would have failed to achieve political eminence', and that you happen to believe for sound, though not generally known, reasons of your own, that 'If Benn had failed to achieve political eminence, he would have been a market gardener in Lowestoft'. Manifestly, you are not at liberty to conjoin and file, as it were, the two propositions, without qualification. (p. 184)

Wright goes on a little later to say:

> I stick to it that the right explanation of such cases is that the transitivity step is valid, but leads to an unacceptable consequence; and, consequently, that no 'filing' may reasonably take place until the premises have been appropriately reformulated or the grounds for their assertion reassessed. (ibid.)

Wright, however, sees this sort of explanation as requiring an acceptance of the claims of his previous paper that conditionals, including counterfactuals, 'sustain transitivity [only] in context' and that what this shows is that conditions are semantically context-dependent (see his 'Keeping track of Nozick', ANALYSIS 43.3, June 1983, pp. 134–40). As he now puts it, what he wants to suggest is that

> The right account of ... conditionals ... may very well involve attributing to them something like, though not to be identified with, token reflexivity: their truth-conditions, that is to say, will be a function both of a context-independent sense and of an envisaged context of utterance. ('Comment on Lowe', p. 184)

In my earlier article, I expressed a sympathy with Wright's desire to defend transitivity, but also an anxiety that his specific proposals regarding the semantics of conditionals threatened to introduce complexities into the logic of conditionals which it were better to avoid if possible. (It is perhaps worth observing that proposals very much like Wright's have also been made recently by Donald Nute, in his 'Logical Relations', *Philosophical Studies* 46, 1, July 1984, pp. 41–56. Compare also Bas van Fraassen, 'Essences and Laws of Nature', in Richard Healey (ed.), *Reduction, Time and Reality*, Cambridge University Press, 1981; though van Fraassen, while advocating context-dependency, still assumes non-transitivity (see p. 198).) I now feel much better disposed towards the idea that an appeal to certain pragmatic features of our use of conditionals can be employed to explain the sort of case Wright illustrates in his new example. However, it is still far from perfectly clear to me exactly what Wright's own position *is*, partly because of his appeal to an analogy with token reflexivity which is nowhere spelled out. For this reason I shall now offer my own (and I hope clearer) explanation of Wright's new example, which I issue without purporting to know to what extent Wright would be prepared to recognize in it just a restatement of his own position.

The explanation I am now inclined to favour is one regarding which I can make no great claim to originality, since its leading idea was presented, in characteristically succinct style, by J. L. Mackie in his brief article 'The transitivity of counterfactuals and causation' (ANALYSIS 40.1, January 1980, pp. 53–4). The idea is that when in ordinary conversation one asserts a conditional of the form 'If p, then q', one implicitly qualifies the antecedent in a fashion which, if it were to be made explicit, would result in one's asserting a conditional of some such form as 'If p *in the easiest way*, then q'. (It is a matter for further debate, which I do not propose to go into here, just how 'easiest' (Mackie's word) is best interpreted – whether in terms of probability or of 'closeness' of possible worlds or indeed in different terms in different cases.) Now, as Mackie points out, even granted the transitivity of conditionals, from 'If p in the easiest way, then q' and 'If q in the easiest way, then r' we cannot validly infer 'If p in the easiest way, then r'. However, if that were all there were to the matter, we should not have succeeded in accounting for Wright's point that, in the sort of case he illustrates, we are 'not at liberty to conjoin and file, as it were, the two propositions [i.e. the two conditionals in his new example], without qualification'. The full explanation is, I think, just this. Precisely because conditionals are transitive and yet when asserted in ordinary conversation carry an implicit qualification to their antecedents, when we actually draw inferences by transitivity in ordinary conversation (as we often do) we automatically adjust the way in which we implicitly qualify the antecedent of our second premise, so as to avoid committing a fallacy of equivocation. If the premises and conclusion of such an inference were made fully explicit, its form would come out as being something like this:

> If p in the easiest way, then q; and if q in *that* way, then r; so if p in the easiest way, then r.

What I am saying, then, is that when in ordinary conversation we assert a conditional *as the second premise of an inference by transitivity*, we do not implicitly qualify its antecedent in the way we normally would when asserting that conditional on its own. Because of this, tension can arise in a case like Wright's, in which two conditionals are each implicitly qualified in the way appropriate for assertion on their own and yet are such that they could, *if the second were implicitly qualified in a different way*, figure in an ordinary conversational inference by transitivity which would lead to an unacceptable conclusion. We need to signal our disavowal of any commitment to the undesirable conclusion without impugning the validity of inference by transitivity; and this can only be done by making more explicit the implicit qualifications that are being made, so as to bring out the fact that we do not really have here the premises required for an inference by transitivity to the undesired conclusion.

If the foregoing suggestions are correct, it means that theorists like Stalnaker and Lewis have made the error of incorporating into their accounts of the *semantics* of conditionals certain features which belong properly to an account of the *pragmatics* of conditionals. Rather than saying, with them, that 'If p, then q' *means* something like 'In the nearest possible world in which p, there also q', we ought it seems to say that there exists a conversational convention to the effect that 'If p, then q' will normally be taken to express what could be expressed more explicitly by something like 'If p in the nearest possible way, then q'. Here it may be objected that no distinct proposition is clearly individuated by the antecedent of the latter conditional. But to this it may be replied that this is just as it should be, since when one asserts a conditional of the form 'If p, then q' in ordinary conversation, one commonly does *not* have a clear-cut and fully articulable antecedent condition in mind; at most one has a rough idea as to what would and what would not be involved over and above what is expressed in p itself (an exception is provided by the case in which the conditional 'If p, then q' is asserted on the grounds that p *entails* q). In asserting an ordinary-language conditional, then, one *indicates* (with more or less precision) but does not *express* the proposition on whose obtaining the consequent is said to be conditional. (If this is the sense in which Wright sees an analogy with token reflexivity, I think I may be in full agreement with him.) Of course, having said just this much still leaves open the question of what *positive* account of the semantics of conditionals one should endorse; but that is a question for a different occasion.

University of Durham,
50 Old Elvet,
Durham DH1 3HN

© E. J. LOWE 1985

A QUICK READ IS A WRONG WRIGHT

By HILARY PUTMAN

STEPHEN READ and Crispin Wright attribute to me a proposal for dealing with sorites paradoxes that I have never made; one which they elegantly demolish (cf. my 'Vagueness and Alternative Logic', *Realism and Reason*, Cambridge 1983, pp. 285-6, and Stephen Read & Crispin Wright, 'Hairier than Putnam Thought', ANALYSIS 45.1, January 1985, pp. 56-8). I find the proposal interesting and the demolition job instructive, but the proposal isn't mine.

They consider a proposal ('Putnam's Proposal') which involves denying:

(1) $(\exists n)(\mathrm{BM}(n) \,\&\, {\sim}\mathrm{BM}(n + 1))$

and which rejects the equivalence of (1)'s denial to

(2) $(n)(\mathrm{BM}(n) \rightarrow \mathrm{BM}(n + 1))$

by adopting intuitionist logic.

They are right that I proposed to use intuitionist logic to block the inference from the negation of (2) – their (5) – to (1). But accepting the negation of (2), which we must do, I think, while *not* inferring (1), is not the same as *denying* (1). In my proposal, (1) is not denied. (What the Read-Wright argument does show, in fact, is that I should deny (1)'s denial – that is, I should accept the *double negation* of (1). This is in keeping with a suggestion I *did* make, that one can use the double negation of a vague sentence to express one's attitude to it.)

Read and Wright do give an independent argument to show that 'intuitionist semantics for negation, appropriately generalized to handle non-mathematical contexts, could [not] fail to endorse the negation of (1)'. This is of independent interest, but it has nothing to do with 'Putnam's Proposal'. For what it's worth, my view about this suggestion is that 'appropriately generalizing' a semantics designed for pure mathematics to 'handle' non-mathematical contexts is not at all straightforward, and the principle that, if we would never accept a statement (and we *know* this) then we *must assert its negation* is *not* one that could be carried over from intuitionist mathematics to arbitrary 'non-mathematical contexts'. In any case, I was proposing a *logic* for inferences involving vague predicates, *not* a 'semantics', so the attack on intuitionist semantics is another example of *ignoratio elenchi.*

Harvard University,
Emerson Hall,
Cambridge, Massachusetts 02138, U.S.A.

© HILARY PUTNAM 1985

DISTRIBUTION, SUPERPOSITION AND QUANTUM LOGIC

By JONATHAN HARRISON

IF a way could be found of saving quantum mechanics from having to reject the distributive law it should surely be welcome. In my paper 'Against Quantum Logic' (ANALYSIS 43.2, March 1983, discussed by Messrs. Gibbins and Hallam in ANALYSIS 44.2, March 1984) I argue that quantum mechanics does not entail rejecting the distributive law. The distributive law has been abandoned because

(1) P0·(V0 v V1...)

has been supposed to be true and

(2) (P0·V0) v (P0·V1) ...

(which is believed to follow from (1) by means of the distributive law) to be false. In fact, however, (1) is as false as (2) is, if quantum theory is true. This, I argued, is because if a particle has an *indeterminate* velocity it must be false that it has any of the *determinate* velocities mentioned in (1)'s second conjunct. I suggested that it may have seemed to many philosophers (Putnam, for example) that the second conjunct of (1) is a logical truth because it mentions all the possible determinate velocities there are. It is not a logical truth, however, but is indeed false, if, over and above the possibility of a particle's having any of these determinate velocities, there is another possibility. This is that it has an indeterminate velocity somewhere or other within the range of velocities mentioned by the proposition V0 v V1..., without determinately having any one of them. This, I thought, must be the possibility that obtained if quantum theory is true. If a particle can have only an indeterminate velocity, the second conjunct of (1) must be false, and so (1) itself false.

It would seem to be natural and convenient to represent this extra possibility, that of a particle's having an indeterminate velocity within (for example) the range of velocities V0, V1, etc., without having any of them determinately by S(V0 v V1 v ...), as it is natural to represent the possibility of truth's being distributed indeterminately over a proposition (perhaps a proposition about the future) and its negation, without being possessed determinately by either severally, by (p v ~p)T. (p v ~p)T, however, implies neither that it is false that p is true, nor that it is false that ~p is true. S(V0 v V1 v ...), however, must be taken to imply that S(V0), S(V1) etc., are all false. S(V0 v V1 v ...) is a *proposition* stating that a particle (S) has a disjunctive velocity (V0 v V1 ...) and so is an atomic proposition, not a disjunctive one. It is important not to confuse 'disjunctions' or 'conjunctions' of positions or velocities, or whatever, with disjunctions or conjunctions of propositions about positions and velocities. I suggest the former should be represented by P0 v P1, P0 · P1, P0 · (V0 v V1), (P0 · V0) v (P0 · V1), etc., the latter by S(P0) v S(P1), S(P0) · S(P1), S(V0) v S(V1), S(V0) v S(V1), etc.

Observing this distinction should show that the considerations about Hilbert space adduced by Hallam are irrelevant to the question whether quantum mechanics involves a rejection of the distributive law. For it is sub-spaces of Hilbert space that are 'conjoined' or 'disjoined', not propositions about them, and there is no breach of the distributive law about *propositions* involved in saying that the sub-space which is the intersection (the analogue of conjunction) of space a with the span (the analogue of disjunction) of space b and space c (i.e., a · (b v c)) is a distinct *sub-space* from that which is the span of the intersection of space a with space b and the intersection of space a with space c ((a · b) v (a · c)). (a · (b v c) = (a · b) v (a · c), which is false, *can* be called a special case of a wider distributive law if 'a', 'b' and 'c' do not have to take propositions as their values, and '·' 'v' and '=' do not have to be interpreted as conjunction and disjunction and equivalence respectively.) The fact that this wide distributive law is false, as it undoubtedly is, does not show that the distributive law about propositions is false, and this is what is at issue. The notation I have suggested makes it absolutely clear that the fact that S(a · (b v c)) does not imply S((a · b) v (a · c)) (a fact which results from the nature of Hilbert space) is not a breach of the distributive law about propositions, for again these are atomic *propositions* about 'conjunctions' and 'disjunctions' of *spaces*, and so the distributive law does not apply to them. (The trouble with V0 v V1 v . . . as a way of representing the velocity of a particle is that it does not reveal the finer structure of V0, etc., as being that of propositions about velocities, and so makes it impossible to bring out the difference between a disjunctive proposition and a proposition about a disjunction of spaces.)

If we represent the fact that a particle has an indeterminate velocity by S(V0 v V1 v . . .), then (1) should be re-written as

(1a) S(P0) · S(V0 v V1 v . . .)

which states that a particle has the determinate position P0 and an indeterminate velocity within the range V0, V1, . . . , i.e., has a velocity which is somewhere within the range, without having any particular determinate velocity within it. (1a) does not imply (2) by the distributive law, for, since (1a) is not a conjunctive proposition with a disjunctive second conjunct, but a conjunctive proposition, whose second conjunct is a proposition about a disjunctive velocity, the distributive law simply does not apply to it. Not only does (1a) not imply (2), it also does not imply the proposition we will obtain when we make the corresponding alterations to (2), viz.,

(2a) S(P0 · V0) v S(P0 · V1) v . . .

For (1a) is a statement to the effect that a particle has a determinate position P0 and an indeterminate velocity, and this cannot imply that it has any of the determinate position-velocities P0 · V0, P0 · V1, etc. It is, indeed, incompatible with its having any of them.

Some problems about superposition can also be avoided by treating the superposition of two states as a disjunctive *state*. The standard way, assumed by Gibbins for example, of dealing with superposition is to describe superposition by means of a disjunctive proposition each disjunct of which states that the particle is in one of the two states superposed. The distinctiveness of superposition is then brought out by saying that though this disjunctive proposition is true, its disjuncts are not true. Since no sane person would wish to maintain that these disjuncts are false this means that they must take neuter truth value. It also means abandoning the rule that it is inconsistent to assert a proposition and to deny that that proposition is true. (This is because from (p v q)T, p ⊃ (p)T, q ⊃ (q)T, you can derive (p)T v (q)T by means of a complex constructive dilemma, and so you must deny p ⊃ (p)T, in order to maintain that a disjunctive proposition can be true, although its disjuncts are not true. (I produce reasons for thinking that it is senseless to deny that the rule p ⊃ (p)T, and a diagnosis of why some philosophers have wanted to reject it, in 'Mackie's Moral Scepticism', *Philosophy* 1982).

If, however, as I suggest, you assert that there is a third, indeterminate state, over and above a particle's being in each of the two states superposed severally, it is possible to avoid having a disjunctive proposition neither (or none) of whose disjuncts is true, and the other anomalies which result from this. This is because if a particle is in this disjunctive state (1 v 2) it is just false that it is in either of the two states 1 and 2 severally, i.e. S(1) and S(2) are both false. There is no logical passage from S(1 v 2) to S(1) v S(2), because the former is not a disjunctive proposition, but an elementary proposition asserting that a proposition is in a disjunctive state, and so S(1) and S(2) are not its conjuncts. Hence there is no need for neuter truth values, and no need to deny the rule that p ⊃ (p)T.

It may be argued, against what I have said, that if states 1 and 2 exhaust the possibilities, as V0, V1, etc. exhaust the possibilities, there is no possibility of a particle's being in a state 1 v 2 which is distinct from either. To this it may be replied that either state 1 and state 2 exhaust the possibilities or they do not. If they do, the particle must be in one or other of these states, which means that it must be true that it is in state 1 or true that it is in state 2. If they do not, there is no reason why S(1 v 2) should not be true, but S(1) and S(2) both false. (My remarks will also apply to route, decay, spin and whatever else it is supposed by quantum physicists to be indeterminate.)

My suggestion opens an escape from a difficulty for quantum mechanics raised by Schrödinger. Schrödinger imagines that a cat is wired to an apparatus in such a way that she will be killed if a particle decays within the hour, not killed if it does not. What, then, happens to the unfortunate animal if the particle neither decays nor does not decay?

If what I have suggested is correct, there is no difficulty. Either the particle's decaying and its not decaying are exhaustive possibilities or they are not. If they are exhaustive, there is no reason in principle why an apparatus should not be produced such that its decaying is a necessary and sufficient condition of the cat's being dead and its not decaying is a necessary and sufficient condition of the cat's being alive. But if, as I suggest, that disjunctive state which is the particle's either-decaying-or-not-decaying is a third possibility, over and above the more normal two, then these other two do *not* exhaust the possibilities. Hence it is impossible to produce an apparatus which is such that the particle's decaying is a necessary and sufficient condition of the cat's being dead, and the particle's not decaying is a necessary and sufficient condition of the cat's being alive. For, if it were in neither of these states, but the third indeterminate state, it would follow that the cat would not be dead, and also that the cat would not be alive, which is impossible. It is, indeed, a logical truth that if two alternatives are each necessary and sufficient conditions for two exhaustive alternatives, they themselves must be exhaustive alternatives. Hence, if decaying and not decaying do not exhaust the possibilities, there is no way of making a machine such that they are necessary and sufficient conditions for two alternatives (a cat's being dead and being alive) which do exhaust the possibilities. This, however, does not show the impossibility of the state, but the impossibility of the machine.

The University,
Nottingham NG7 2RD

© JONATHAN HARRISON 1985

WEIR AND THOSE 'DISPROOFS' I SAW BEFORE ME

By NEIL TENNANT

IN his *Rejoinder to Tennant* (ANALYSIS 45.2, March 1985) Weir complains that the anti-realist owes him an account of why it is justified to move from denying that a speaker does not understand a sentence to asserting that he does. The anti-realist's reply must be that the question whether a speaker understands a sentence or a particular word or a particular kind of construction is one that can always be decided after appropriate investigation. This not to say that at any stage of one's acquaintance with the other speaker's use of the language, one is able definitely to pronounce on his competence one way or the other. It is merely to say that, should the question of his competence in any regard ever be raised, we know what the appropriate avenues of investigation are that would very quickly put us in a position confidently to be able to assert that he does after all understand the word or construction in question or confidently to be able to deny the claim. This is why the classical rule of dilemma:

$$\begin{array}{cc} \overline{A}^{\,(i)} & \overline{\sim A}^{\,(i)} \\ \vdots & \vdots \\ B & B \\ \hline \multicolumn{2}{c}{B} \end{array}\;(i)$$

is justified when the assumptions A and ~A for the two case proofs are, respectively, attribution of the understanding in question and its denial. Attributions of semantic understanding are a special kind of psychological report. They happen to be decidable.

It is in examples such as *attributions* of understanding, and other psychological reports, that Weir forces the discussion away from the straightforward cases of claims with the nature of whose understanding I was concerned. He claims that this is quite in order, since he is concerned to bring out a tension between maintaining determinacy of meaning and rejecting determinacy of truth value. I confess to feeling no such tension. My example was of a mathematician's understanding of an arithmetical statement. In such a case, the means for deciding a statement one way or the other are more strictly circumscribed than they are in the case of a psychological report. One is better placed to see in advance what the general shape of conclusive evidence for one's assertion would be. To say this is not to lose sight of the fact that the notion of proof, for the anti-realist in mathematics, is still a reasonably open-textured one. The reason why I am at a loss to understand why possession of appropriate proof-checking abilities should not be course would have to be an anti-realist specifically on subjunctives. This may not, however, be so.

A further cautionary note should be entered on the locution 'being an anti-realist on ...'. I don't, of course, intend to suggest that we are free to take up or drop the anti-realist's basic postulates of publicity and compositionality in a given area of discourse simply as the spirit moves us. These two postulates, rather, hold across the board and are the starting points of any anti-realist critique. I assume them as obvious for any area of discourse. Rather, 'being an anti-realist on' any given area of discourse is arrestingly different from the classical position only when the *result* of prosecuting those principles to the full is a refusal to accept orthodox moves in reasoning. An example would be a refusal to regard every statement in the area as decidable, and hence a refusal to apply to it the law of excluded middle or the rule of dilemma or any of the usual classical equivalents. Whether one should reach such a result in the case of subjunctives is a matter I cannot discuss in detail here.

Weir believes himself able to

> assuage the self-doubts of realists that the classical conception of quantifiers, as exposing infinitistic truth functions, and hence capable of generating undecidable sentences, *clashes with* plausible views about how we could grasp such truth functions. (my emphasis)

To assuage *that* self-doubt is easy. More difficult, however, is to assuage the doubt that the anti-realist is more subtle in entering: to wit, that the classical conception *goes unjustifiably beyond* plausible views about how we could grasp such truth functions.

Once again, this is where a deep misunderstanding of the anti-realist's position surfaces in the writings of a realist. It is the extra theoretical slack in the realist's account, the slack that has no behavioural import on any methodology sensitive to normative issues, that the anti-realist is concerned to eliminate.

If this 'revolution' involves abandonment of classical logic for intuitionistic logic, or indeed any other, we have yet to see whether Weir's misgiving is correct that this would have a 'debilitating effect ... on our reasoning, our mathematics and, through it, on our physical science'. It is a major concern of mine in [198?] to show that this is emphatically not so. I can summarize here results already obtained for the system of intuitionistic relevant logic, which I would argue for as the logic that is justified on an anti-realist account of the meanings of the logical operators: It is adequate for the development of a theory of truth adequate in Tarski's sense. It is adequate for the falsificationist method in science. It is adequate for the foundations of constructive mathematics. It is therefore up to Weir to substantiate his misgivings at being deprived of strictly classical canons of reasoning.

The University of Stirling,
Stirling FK9 4LA

© NEIL TENNANT 1985

REFERENCES

1977 H. Field, 'Logic, meaning and conceptual role', *Journal of Philosophy*, 74, pp. 379-409.

1981 N. Tennant, 'Is this a proof I see before me?', *Analysis*, 41.3 (June 1981) 115-19.

1983 H. Leblanc and C. G. Morgan, 'Probabilistic semantics for intuitionistic logic', *Notre Dame Journal of Formal Logic*, 24, pp. 161-180.

1983 A. Weir, 'Truth conditions and truth values', *Analysis*, 43.4 (October 1984) 176-80.

1984 N. Tennant, 'Minimal logic is adequate for Popperian science' (forthcoming in the *British Journal for Philosophy of Science*).

1984 N. Tennant, 'Were those disproofs I saw before me?' *Analysis*, 44.3 (June 1984) 97-105.

1985 A. Weir, 'Rejoinder to Tennant', *Analysis*, 45.2 (March 1985) 68-72.

198? N. Tennant, *Anti-Realism and Logic*, Oxford.

NOONAN, 'BEST CANDIDATE' THEORIES AND THE SHIP OF THESEUS

By B. J. GARRETT

IN 'Wiggins, artefact identify and "best candidate" theories' (ANALYSIS 45.1, pp. 4-8), Harold Noonan discusses the 'best candidate' theorist's description of three possible histories of the Ship of Theseus and poses a dilemma for such a theorist. I wish to argue, however, that this dilemma can be resisted and that either horn can be accepted without impalement.

The three possible situations involving the Ship of Theseus can be pictured as follows:

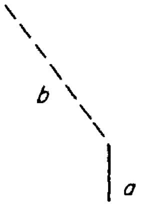

Figure 1

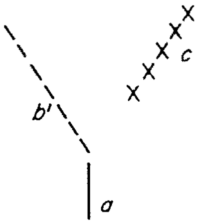

Figure 2

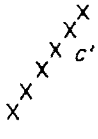

Figure 3

In all three situations 'a' denotes the Ship of Theseus. In situation 1, 'b' denotes the continuously repaired ship. In situation 2 (the situation described by Hobbes), 'b'' denotes the continuously repaired ship and 'c' denotes the plank-hoarder's ship, reconstituted from a's original planks. In situation 3 the ship reconstituted from a's original planks is denoted by 'c''. Given that we regard spatio-temporal continuity (under a sortal) as the most important criterion of artefact identity, outweighing the identity of original parts criterion, but acknowledge the fact that an artefact can survive dismantlement and reassembly, the most natural description of the above situations is as follows: in situation 1, a = b; in situation 2, a = b′ and a ≠ c; and in situation 3, a = c′.[1] This commonsense description coincides with that of the 'best candidate' theorist: the continuously repaired ship and the plank-hoarder's ship are both *candidates* for identity with the Ship of Theseus, though the former ship has the stronger claim.

However, Noonan presents the following dilemma to the 'best candidate' theorist: the theorist must hold either that c = c′ or that c ≠ c′. Whichever he chooses, claims Noonan, the result is absurdity. Suppose that the 'best candidate' theorist identifies c with c′. Then, endorsing an argument of Wiggins, Noonan claims that we could truly say of ship c in situation 2 that, but for the existence of its rival, b′, *it* would have been the Ship of Theseus – and this is absurd.

Noonan assumes that if the 'best candidate' theorist identifies c with c′, he must jettison the necessity of identity. This is true only if identity is a transitive relation. However, it has been suggested by some philosophers (e.g. Arthur Prior in 'Time, Existence and Identity', *Proceedings of the Aristotelian Society*, 1966, pp. 183–92) that we ought to regard the 'multiplication by division' of unicellular organisms, and the possible division of conscious organisms, as demonstrating that identity is non-transitive. For, claims Prior, the correct description of cases of division is that one thing is identical with two things which are themselves distinct, and this description of such cases is available to a 'best candidate' theorist. It is quite consistent for such a theorist to hold that in cases where two candidates possess equally sufficiently strong claims for identity with the original object, the original is identical with both candidates *and* that in cases where the candidates possess claims of unequal strength, the stronger candidate is identical with the original.

A further reason for regarding identity as non-transitive is this. It might be claimed that it is possible for a material object to undergo a series of changes, each of which is identity-preserving, yet the total effect of which is to yield a non-identity. For example, suppose that a table A is gradually transformed, molecule by molecule, into a chair Z. At the end of the process A no longer exists: A is not the same artefact as Z. (The alternative description – A still exists but is now a chair – is implausible.) Yet, given that the relation 'same artefact' is *tolerant* with respect to the concept

---

[1] Though not always. If, in a type-2 situation, the planks are removed and replaced too quickly, the original ship may be considered to be identical with the plank-hoarder's ship. (Cf. Nathan Salmon, *Reference and Essence* Oxford 1982, p. 221.) If so, the continuity criterion requires a temporal qualification if it is to provide a sufficient condition for the identity of artefacts over time.

number of molecules (e.g. that changes of just one molecule cannot affect the identity of an artefact), it follows, where 'B', 'C', 'D' etc. denote the artefacts present after each successive change, that the identity statements $A = B$, $B = C$, $C = D, \ldots, Y = Z$, are all true. But $A \neq Z$; hence identity is non-transitive. Relinquishing the transitivity of identity may thus yield a solution to certain versions of the Sorites Paradox.

If, for such reasons, the 'best candidate' theorist regards identity as non-transitive, he can identify c with c′ consistently with the necessity of identity. It will then not be true to say that, but for the existence of b′ in situation 2, c would have been the Ship of Theseus, since situation 3 is not a situation in which c is identical with the Ship of Theseus. A 'best candidate' theorist who takes this line can identify c with c′ without incurring Noonan's charge of absurdity.

However, Noonan points out, *contra* Wiggins, that the 'best candidate' theorist need not identify c with c′. But, claims Noonan, this denial commits the 'best candidate' theorist to the following 'self-evident' absurdity: if $c \neq c'$, then we can truly say that, but for the existence of b′ in situation 2, c would not have existed *at all*. (Note that the 'best candidate' theorist cannot slip through the horns of Noonan's dilemma by claiming that '$c = c'$' is *indeterminate* in truth-value. For Noonan's charge would still hold: if it is indeterminate that $c = c'$, then we can truly say that, but for the existence of b′ in situation 2, it would have been indeterminate whether c existed, and this is absurd.) However, a 'best candidate' theorist can justifiably deny that this consequence is absurd. For although it is true that, but for the existence of b′, c would not have existed, this fact is not important given that a ship, materially indiscernible from c, would then have existed, which stands in the same material relation to the Ship of Theseus as c does in situation 2. (Derek Parfit has made the same reply informally to a similar objection to Reductionist theories of personal identity: for since, according to Parfit, 'what matters' in ordinary survival is not identity, and hence not existence, the fact that, if I am one of the products of a successful fission, had the other 'off-shoot' not taken, *I* would not have existed, is not important given that a person would then have existed who is psychologically, and physically, identical with me, and who is psychologically continuous with the pre-fission person.) Thus if the 'best candidate' theorist takes the Parfitian view that identity, and hence existence, is not 'what matters', the fact that, but for the existence of b′, c would not have existed at all, ceases to be a significant fact, and hence ceases to be an absurd one.

In this reply my aim has not been to endorse one or other of the above positions but merely to indicate ways in which a 'best candidate' theorist might respond to Noonan's dilemma. It is important to realise that Noonan's argument must be defused since it poses a threat, not just to 'best candidate' theories of identity, but to the tenability of the common-sense description of situations 1-3. If the choice is between rejecting that common-sense description and endorsing one of the two positions outlined above, the least unpalatable option is, I think, the latter.

Lincoln College, Oxford © B. J. GARRETT 1985

REPLY TO VAN BRAKEL

By D. K. BUCKNER

MY criticism of the 'demonstrative' theory of quotation ('Goldstein on Quotation', ANALYSIS 44.4, October 1984, pp. 189-90, commenting upon L. Goldstein's 'Quotation of Types of Other Types of Quotation', ANALYSIS 44.1, January 1984, pp. 1-5) consisted of the following argument. We cannot assimilate the '"Kuznetzin"' of

(1) The word 'Kuznetzin' denotes the former Russian ambassador

to the ': Kuznetzin:' of

(2) This word: Kuznetzin: denotes the former Russian ambassador

since, though both sentences mean the same, the word 'Kuznetzin' is read as a part of the sentence in (1), but not in (2). *And since it is fundamental to the Davidson theory that we do not read "quoted" expressions*, it follows that the inverted commas in (1) are not Davidson quotes, nor is the 'Kuznetzin' of (1) a Davidson-quoted expression.

J. van Brakel ('Buckner Quoting Goldstein and Davidson on Quotation', ANALYSIS 45.2, March 1985, pp. 73-5) objects that the premiss italicised above is 'a gross misinterpretation of the demonstrative account' since 'nowhere is it stated, by either Goldstein or Davidson, that we do not read the "quoted" sample'. He is right that neither actually says this, but it is nonetheless implicit in the stated versions of the theory. According to Goldstein (p. 4) quotation marks (Davidson marks) are like 'arrows pointing to a sample which could be remote from the sentence'. Well, suppose I produce a spoken delivery of (2) at a lecture. Then I utter 'This word denotes the former Russian ambassador' while pointing with a stick to a word on the blackboard. The stick is a Goldstein 'arrow'; so are the double-colons in (2); hence the double-colons are genuine examples of Davidson-marks. But 'Kuznetzin' is not read in (2), and it seems that, ex hypothesi, we do not read the contents of Davidson-marks. Again, Davidson suggests (‘Quotation’, *Theory and Decision* 11 (1979), pp. 27–40 p. 38) that a Davidson-quoted expression is analogous to the fish I hold up while uttering ‘I caught this fish today’. As he rightly says, the fish is not a part of the language; nor is the ‘Kuznetzin’ pointed to on the blackboard; nor, as we all agree, is the ‘Kuznetzin’ of (2). It is surely intrinsic to the demonstrative account that a sample that is “quoted” in the Davidson manner does not function as part of the language. Hence the inverted commas of (1) cannot be Davidson-style quotes.

Now van Brakel objects that ‘quotation marks can always be eliminated by using some other conventions that serve the same goal’. Quite so: (1) has pretty much the same sense as (2). But does it follow from this that the quotation marks in (1) are just Davidson-marks in another guise? Not at all: from the fact that the sequence of marks ‘the word ‘Kuznetzin’’ behaves, in combination with ‘denotes the former Russian ambassador’, just like the sequence ‘this word: Kuznetzin:’ we are not entitled to infer that the component parts of each sequence operate in the same way. My point, which is simple enough, is that there must be *something* wrong with the Davidson-Goldstein account if it ignores the important and obvious distinction between (1) and (2).

Nor is this objection so trivial as van Brakel apparently suggests, once it is considered in the proper context. Goldstein's original point (‘Quotation of Types . . .’) was that one should *not* consider the ‘Kuznetzin’[1] of (1) as a sort of name-in-apposition like the ‘E♭’ in

(3) The note E♭ begins *Rheingold*

But of course the difference between (1) and (2) is precisely that which distinguishes (3) from

(4) This note (hum) begins *Rheingold*

The humming sound that would accompany my utterance of (4) is as grammatically irrelevant as the ‘Kuznetzin’ of (2). By contrast, one cannot remove the ‘E♭’ of (3) without drastically altering the sense of the whole sentence; similarly for the ‘‘Kuznetzin’’ of (1). The suggestion that non-Davidsonian quotation-expressions *do* function as names-in-apposition is not one that the demonstrative theory of quotation can dispel.

I will readily admit that Goldstein's difficulty is a real one. If the ‘Kuznetzin’ of (1) functions (with the aid of quotation marks) as a sort of name, then what does it name? The token that (accidentally) occurs between quotation marks? Or the type of which that token is a token? But the same difficulty attaches to the 'E♭' of (3). Does it name the note that begins *Rheingold*? Or the fourth note of Beethoven's *Fifth*? Which performance of that note is intended? What, for that matter, does '*Rheingold*' name? The problem is apparent enough. But it is not one with which the Davidson-Goldstein theory can satisfactorily engage.

---

[1] or ‘‘Kuznetzin’’: there is a further points, which I shall not consider here, as to whether the inverted commas should be included as a part of the quotation-name (so that ‘‘Kuznetzin’’ operates in *suppositio formalis*), or whether they indicate a change of reference (making ‘Kuznetzin’ work in *suppositio materialis*.)

University of Bristol,
9, Woodland Road,
Bristol BS8 1TB

© D. K. BUCKNER 1985

GREEN AND DICTATORS REVISITED

By DAVID GORDON

LESLIE Green argues that it is not obvious that an economically rational person '... would regard a procedure which did not allow others to vote as being the best.... [It is not] obvious, but dubious, that one would have to be irrational not to want to govern others as well.' ('What Is A Dictator?' ANALYSIS 45.2, March 1985, p. 127.)

This seems to me perfectly correct. But it is not, as Green thinks, a point against me. In my previous criticism of him, I maintained neither that an economically rational person would regard a procedure in which others were not allowed to vote as best or that this view is an obvious truth. Rather, I claimed it to be obvious that 'there is *best reason* for someone to participate in a social choice if he thinks his preference will be absolutely decisive'. ('Green on Dictators and Democracies', ANALYSIS 44.2, March 1984, p. 96.)

It does not follow from my claim that it is irrational not to prefer a situation in which others are prevented from voting to one in which they are not. It also does not follow that an economically rational person must prefer a situation in which no one other than himself in fact shows up to vote, though there is no prohibition in force upon anyone's voting, to that in which most people as a matter of fact do vote. My argument was solely about when one has best reason to vote: I advanced no claim as to what system of election it is rational to prefer.

Suppose, for example, that someone thinks he has best reason to vote in an election for president of a club if he believes no one but himself will appear at the election meeting. He need not favour a change in the club rules forbidding anyone else from casting a vote. He is also not inconsistent in hoping that everyone in the club turns out for the meeting. (Perhaps he thinks this will be good for the club and values the club's interests more than his having the decisive vote in electing the president.) Green has wrongfully conflated the rationality of voting with preference for a particular system of election.

The new difficulties Green raises for the non-dictatorship requirement (D) used in social choice theory also seem to me to be based upon misunderstanding. He invents three figures, Bored, Scared, and Lucky, each of whom he contends counts as a dictator under D, although it is strongly counterintuitive to regard any of them as in fact a dictator.

But are Bored, Scared, and Lucky dictators under D? I do not think so. According to D, '"There is no individual whose preferences are automatically society's preferences independent of the preferences of all other individuals."' (p. 125, quoting K. J. Arrow, 'Values and Collective Decision-Making') Scared and Lucky do not have preferences which are automatically society's independent of the preferences of all other individuals. Instead, if their preferences happen to coincide with the majority's, they will be on the winning side. (Similarly, so long as Bored is indifferent to all social choices, he will not find the social choice against his wishes.) Should the preferences of Bored, Scared, or Lucky cease to be tied to those of others in the way Green sets forward in his definitions, they would cease to be necessarily those of society. Bored, Scared and Lucky's preferences are not, as D requires, the preferences of society independent of the preferences of all other individuals: they are so only if connected with the preferences of other individuals in a definite fashion.

Green is also incorrect to think that if social choice theory relies 'only on the calculus of preference relations', it cannot exclude Bored, Scared, and Lucky from qualifying as dictators under D except by 'imposing *ad hoc* restrictions on preference orderings ... but that is to give up the condition of unrestricted domains on which Arrow's theorem depends'. (p. 126) On the contrary, if one uses Condition 1′ of Arrow's *Social Choice and Individual Values*, viz., 'All logically possible orderings of the alternative social states are admissible' (2nd edition, New York, 1963, p. 96), it is easy to see how Bored, Scared, and Lucky can be excluded as dictators without the use of *ad hoc* devices. Unless it is claimed to be *logically necessary* that anyone who is Bored, Scared, or Lucky must have the defining properties of one of the three (and claiming this seems an *ad hoc* device if anything does), there will exist logically possible orderings in which the preferences of Bored, Scared, or Lucky are in the minority. Suppose, e.g., a society of three individuals, Scared and two others, choosing between goods X and Y. It is logically possible that Scared choose X and the other two choose Y. Under majority rule, Y is the social choice. He is thus a non-dictator, and Green's criticism fails. (I do not mean to suggest that Condition 1′ is necessary for the proof of Arrow's

General Possibility Theorem: I intend this only as an example of how Green's point can be met.)

6238 Orange Street,
Los Angeles, California 90048, U.S.A.

© DAVID GORDON 1985

AGNOSTICISM

By THOMAS V. MORRIS

MANY philosophers have believed that there is a simple principle of rationality which properly governs our assessment of positive existence claims in the absence of any grounds for their truth. Versions of such a principle have been stated by a number of writers, including prominently the late Norwood Russell Hanson (see the title essay of his posthumously published *What I Do Not Believe and Other Essays*, edited by Stephen Toulmin and Harry Woolf, Dordrecht 1972), and Michael Scriven (*Primary Philosophy*, New York 1966, p. 103). Updating and refining a bit what they had in mind to accord with recent developments in epistemology, let us consider a formulation of what we can call, in honour of their sponsorship, the Hanson-Scriven Thesis:

> HST – For any rational subject S and any positive existence claim P, if S is in possession of no good evidence or any other positive epistemic ground for thinking that P is true, then S ought to adopt the cognitive relation to P of denial.

In a recent paper entitled 'Necessary Agnosticism?' (ANALYSIS 44.4, October 1984, pp. 198-202), Robert McLaughlin has endorsed such a principle, specifying that, in his words, 'total lack of grounds for an ontically significant proposition is itself a sufficient ground for its negation' (pp. 198, 199). It is McLaughlin's concern to insist that no ontically significant proposition, no positive existence claim, is immune to evaluation by this principle, which he presents as a version of Occam's Razor. Thus he holds, along with Hanson and Scriven, that if a person is considering a positive existence claim, any such claim whatsoever, and finds himself without any good grounds to affirm it, he is rationally required to deny it, affirming its negation instead.

A number of theists, as McLaughlin recognizes, have held otherwise with respect to the proposition that there is a God. The view here is that a person can rationally judge himself to possess no good ground for thinking theism to be true, and yet this alone not justify him in being an atheist. So rather than atheism's being rationally required on the part of anyone merely bereft of any positive grounds for theism, it is not even rationally permitted by this condition alone. The theistic claim is that unless he has sufficient, independent grounds for atheism, agnosticism is required of such a person if he is to be rational.

McLaughlin seems to think that all theists who hold such a view mean to exempt the proposition that there is a God from the scope of any sort of Razor principle. This is, strictly speaking, false. Further, he assumes that the only reason for thinking agnosticism necessary in the mere absence of any positive ground for theism turns on an argument concerning the modal character of the theistic existence claim. I want to suggest that it is not the modal character, but rather the metaphysical character, of the claim that there is a God which should prevent any person with no good reason for thinking it true from, on that ground alone, thinking it false.

McLaughlin identifies a possible line of reasoning from the modal status of the theistic claim to the necessity of agnosticism in what he terms justificatory vacuum. On a standard view of deity, the proposition that God exists has a high modal status: if it is true, it is necessarily true; if false, necessarily false or impossible. The argument then is that a person is rational in denying a claim of high modal status only if he has a proof which shows directly the impossibility of the claim. Thus, no such claim can ever be rationally denied on the sole basis that there exist no positive grounds in its favour. So neither the HST nor any other Razor-style principle can apply to a positive existence claim of high modal status. Atheism is rationally justified only if it is proved directly that the existence of a God is impossible – an extraordinarily difficult task indeed, and one that has not yet been clearly discharged by anyone. Agnosticism is necessary in absence of any direct modal impossibility proof.

McLaughlin has no trouble successfully refuting this line of argument. Existence claims of high modal status can be evaluated by non-modal means, as for example by inductive arguments and appeals to intuition (I have explored the role of intuition concerning such modal claims in 'Necessary Beings', *Mind* 95, April 1985). There is thus no reason in principle why something like the HST cannot apply to an existence claim just because the claim has this modal status. So pointing to its modal character will not suffice to exempt the proposition that there is a God from the scope of a properly formulated non-modal rule for the evaluation of positive existence claims.

But is the HST properly formulated? Does it, or does McLaughlin's shorter version which it reflects, adequately capture a rule we all use, or ought to use, for the rational assessment of positive existence claims? Let us consider a particularly clear example of a case where such a rule seems to be applied.

Suppose that while seated at the desk in my office, the proposition suddenly occurs to me that there is a large poisonous snake in the room. Although this unpleasant thought has entered my mind with no positive lustre of truth about it, and has elicited from me upon its appearance no feeling of conviction at all, it nonetheless has presented itself with sufficient force as not to be ignored. And after all, once entertained seriously, such a disturbing thought is not easily dismissed. So I make a quick but thorough search of the place, which is neither unduly large nor unusually cluttered. During the process, I am not appeared-to snakily, nor do I find any trace of such a visitor. What then should be my cognitive attitude toward the proposition in question, given that I possess absolutely no evidence or any other positive epistemic ground in its favour? Surely that of rejection or denial. The only rational response in such conditions is to believe the proposition to be false. And this judgment accords perfectly with the precise formulation of the Hanson-Scriven Thesis with which we began.

So, many philosophers seem to have concluded from a consideration of such cases as this, and more ordinary ones as well, that in the HST we have a properly formulated rule for the rational assessment of any positive existence claim. But is this correct? In a recent article, Alvin Plantinga has rejected the view of Hanson and Scriven with a very interesting argument (presented in his major paper 'Reason and Belief in God', in *Faith and Rationality*, edited by A. Plantinga and N. Wolterstorff, Note Dame 1983, pp. 27–9). He first asks us to consider the proposition

(1) There is at least one human being that was not created by God.

Since it is, he points out, a necessary, conceptual truth about God (at least about the God of orthodox Judeo-Christian theology) that

(2) If God exists, then God has created all the human beings there are,

any set of epistemic considerations in favour of the truth of (1) – evidences, arguments, etc. – will have to contain an argument that there is no God. But suppose that

(A) There is no good argument against God's existence.

Then we shall have as a consequence no good argument for (1) either, and so, according to Hanson and Scriven, we must believe its denial:

(1′) All human beings are created by God.

Now suppose also, Plantinga suggests, that

(B) There is no good argument for God's existence,

a supposition which surely would be dear to the hearts of philosophers such as Hanson, Scriven and McLaughlin. Then when considering

(3) There is a God

we find that we are obliged to believe its denial

(3′) There is no God.

But then, Plantinga continues, assuming that

(C) There are no good arguments for or against the existence of God,

the Hanson-Scriven view forces us to believe

(4) There is no God and all human beings are created by him

which is awkward and embarrassing enough by itself, but which also, along with the obvious truths that

(5) Some human beings exist

and that

(6) No human being was created by God unless God exists

entails a contradiction. And certainly, no principle which produces this sort of absurdity is a basic rule by which we are and ought to be guided in our rational assessments of positive existence claims. Thus, the HST must be rejected.

However, interesting as this argument is, we can well imagine that Hanson, Scriven, and McLaughlin would be quick to protest that this attempted refutation of their view just begs the question against them. For on their favoured principle, assumption (C) could never be true. The absence of any good argument for the existence of God would just itself provide us with a good argument against the existence of God. And without (C), no absurdity follows.

So, is the HST without flaw? Does it provide us with a basic principle of rational belief allotment? The suggestion of something like it has stood for years in the literature, unrefuted, and apparently endorsed by many philosophers. And it has the apparent backing of many common intuitions about the proper governance of our assent, as illustrated by the story of the poisonous snake. However, it is easy to show that such backing is apparent only for the precise version of the HST we are considering, the one which accords with McLaughlin's remarks.

Circumstances in which the lack of any positive epistemic considerations (evidence, testimony, natural inclination to believe, etc.) for some positive existence claim P rationally oblige a subject S to deny P rather than to withhold judgment on it, or remain agnostic, are those and only those in which S reasonably believes he is in good epistemic position with respect to P, where being in good epistemic position relative to a proposition is being in such a position that should there be any positive epistemic considerations for the truth of the proposition, one would have them, or most likely have them. It is possible to be in good epistemic position with respect to a proposition only if it is such that, should it be true, its truth would be indicated by some positive epistemic considerations in its favour. When I have completed my search for the poisonous snake, I am in good epistemic position to assess the claim that such a thing exists in my office. In that position, and only in that position, lack of any positive epistemic consideration for the claim amounts to as decisive an epistemic consideration against it as one could want. It is only when a person reasonably believes himself to be in good epistemic position for assessing an existence claim that the absence of any evidence or other positive consideration for it can warrant and require his denial of the claim.

So clearly, the HST should be revised to read something like:

> HST′ For any rational subject S and any positive existence claim P, if S reasonably believes himself to be in good epistemic position relative to P, and S is in possession of no good evidence or any other positive epistemic ground for thinking that P is true, then S ought to adopt the cognitive relation to P of denial.

Now, what is important for our purposes and relevant to an evaluation of McLaughlin's views about theism, atheism and agnosticism is the recognition that a theist can endorse HST′ as a principle governing the rational assessment of all positive existence claims, and at the same time deny that, with respect to the claim that there is a God, the mere lack of any grounds for thinking it true can suffice to justify a person in thinking it false. If S finds himself without positive grounds for P, either of two things can be true: (1) P is false, or (2) S is not in good epistemic position relative to P. Because of some interesting features of the proposition that there is a God, it is particularly difficult for a person lacking any positive grounds for theism and any independent grounds for atheism to judge reasonably that he is in good epistemic position relative to this claim.

First of all, the proposition in question is a metaphysical existence claim. As such, it is very different from any such claim as that there is a snake in my room, or that a certain planet has a moon. Almost any major metaphysical view can be expressed in the form of a positive existence claim. And any such existence claim will be just as difficult to assess as any fundamental metaphysical stance ever has been. The problem with metaphysical existence claims is this: if we have a proof that such a claim is true, a proof operating on premises of whose truth we are sure, with argument forms of whose truth-preserving nature we are sure, then we can be sure that we are in good epistemic position with respect to that claim. Likewise, if we have such a disproof. But without such a proof or a disproof, we are often without good grounds for making any reasonable judgment either that we are or that we are not in good epistemic position relative to the claim. Consider any major metaphysical claim, disputed over the centuries. Under what conditions could we find ourselves without any convincing epistemic grounds for the claim, and without independent grounds against the claim, and yet be in a position reasonably to judge that if there were any positive grounds for its truth, we would be in possession of them? How could we even have good reason in such circumstances to think that the proposition in question is such that if it were true, its truth would be manifested in the existence of positive epistemic considerations in its favour? That is, how can we reasonably judge that it would even be possible to be in good epistemic position concerning it? These are questions which would be in most if not all cases exceedingly difficult to answer with any degree of confidence. Surely, in a justificatory vacuum, with respect to a metaphysical existence claim, it can be rational to withhold judgment concerning one's epistemic position and thus concerning the claim itself. Denial of the claim is certainly not required under such conditions by any true principle of rationality. In fact, without any sort of independent grounds for the claim's negation, it can be held that denial would be not only rash but irrational as well. And this is fully consistent with the revised principle HST′.

Since the proposition that there is a God is a metaphysical existence claim, a person can judge himself to lack any positive epistemic ground for its truth, and yet that fact alone neither require nor even justify his being an atheist. And it is now clear that holding this need not issue from any effort to exempt the theistic claim from the scope of a properly formulated Razor-style principle. From its being a metaphysical claim, it does not follow that the HST′ does not apply to the proposition that there is a God, only that this principle will not generate in its case the simple results McLaughlin wants.

One further consideration should be mentioned as well. Suppose a person takes himself to have no good ground for theism and no independent ground for atheism. Either there is a God or there is not (for whatever precise conception of God is most adequate). If God exists, a failure to have any positive considerations in favour of this alternative (a lack of evidence, and of a natural inclination to believe, etc.) may be due to the noetic effects of sin. So many theists claim. And on the other hand, if there is no such being, a failure to see that there is not may be due to low-grade effects of a deep psychological need not to believe this. In either case, one who lacks evidence and other positive considerations could fail to be in good epistemic position, having his noetic abilities clouded in one way or another. Realizing these possibilities alone can rationally warrant agnosticism in these conditions.

With or without this latter point, we can see that McLaughlin is wrong in thinking that a Razor-style principle which governs the rational assessment of positive existence claims clearly requires, in the mere absence of any positive ground for theism, atheism. The epistemic dynamics of religious belief are quite different from that.

The University of Notre Dame,
Notre Dame, Indiana 46556, U.S.A.

© THOMAS V. MORRIS 1985

AUSTRALIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 63, No. 4; December, 1985

EDITOR: BRIAN ELLIS

ARTICLES

B. L. Davidson Belief *de re* and *de se*

George R. Carlson Hume and the Moral Realists

Toomas Karmo Are Singular Terms Needed for Describing the World?

A. E. Pitson Frank Jackson and the Characterisation of Sense-Data

P. J. Crittenden Sartrean Transcendence: Winning and Losing

Michael Wreen Vagueness, Values, and the World/World Wedge

Paul Weirich Decision Instability

DISCUSSION

Ellery Eells Weirich on Decision Instability

Alan E. Fuchs Rationality and Future Desires

C. L. Hardin Frank Talk About the Colours of Sense-Data

Michael Pendlebury Hetherington on Possible Objects

Don Mannison Meaning and Metaphor

CRITICAL NOTICES

Robert Elliot *Ecological Ethics and Politics* by H. J. McCloskey, and *The Ethics of Environmental Concern* by Robin Attfield

Kim Sterelny *From Folk Psychology to Cognitive Science* by S. P. Stich

REVIEW ARTICLE

Maurita Harney The Philosophy of Technology Review of works by P. Durbin and F. Rapp, F. Rapp, and S. Turkle

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$17 (US$18 or Stg£10); Institutions A$34, (US$36 or Stg£18). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

THEORIA

A Swedish journal of philosophy
Edited by Peter Gärdenfors, Bengt Hansson
and Krister Segerberg

Volume 49, part 1
SPECIAL ISSUE ON EVOLUTIONARY EPISTEMOLOGY

A. Olding Biology and knowledge
Andrew J. Clark Meaning and evolutionary epistemology
Neil Tennant In defence of evolutionary epistemology

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Editor,* THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-223 50 Lund, Sweden.* Subscription for one year is SwCr 60:– or U.S. $ 16.–, for three years SwCr 160:– or U.S. $ 40.–.

INQUIRY

Editor: Alastair Hannay

Vol. 28 (1985), No. 2	
Disarming Nuclear Apologists	Robert E. Goodin
Faces	Avrum Stroll
The Closest Continuer Theory of Identity	Harold W. Noonan
Human Nature and Moral Theories	John Kekes
Addis on Analysing Disposition Concepts	Fred Wilson
Hamlet without the Prince of Denmark Revisited: Pörn on Kierkegaard and the Self	Alastair Hannay
Subjects, Intelligibility, and History (Charles Guignon, *Heidegger and the Problem of Knowledge)*	Theodore R. Schatzki
Vol. 28 (1985), No. 3	
Marxism and Radical Democracy	Joseph V. Femia
The Incoherence of Intergenerational Justice	Terence Ball
Hermeneutics and the Social Sciences: A Gadamerian Critique of Rorty	Georgia Warnke
Rationality and Psychological Explanation	John Heil
Reason and Personal Identity (Derek Parfit, *Reasons and Persons)*	Alan H. Goldman
The Fate of the Enlightenment Program (Nicholas Maxwell, *From Knowledge to Wisdom: A Revolution in the Aims and Methods of Science)*	John Kekes

Subscription price:
Institutions: N.kr. 300,- $50.00 — Individuals N.kr. 155,- $26.00.

Published quarterly by Universitetsforlaget, P.O. Box 2959 Tøyen, 0608 Oslo 6, Norway
— P.O. Box 258, Irvington-on-Hudson, New York 10533, USA.

SOCIAL PHILOSOPHY & POLICY

Editor: Ellen Frankel Paul
Associate Editors: Fred D. Miller and Jeffrey Paul

Volume 2, no. 2 March 1985

Ethics and Economics

AMARTYA SEN	The Moral Standing of the Market
ALLAN GIBBARD	What's Morally Special about Free Exchange?
DAVID GAUTHIER	Bargaining and Justice
ALEXANDER ROSENBERG	Prospects for the Elimination of Tastes from Economics and Ethics
JULES L. COLEMAN	Market Contractarianism and the Unanimity Rule
JOHN C. HARSANYI	Rule Utilitarianism, Equality, and Justice
DAVID COPP	Morality, Reason, and Management Science: The Rationale of Cost-Benefit Analysis
E.J. MISHAN	Consistency in the Valuation of Life: A Wild Goose Chase?
DAN USHER	The Value of Life for Decision Making in the Public Sector

Subscription Rates
Volume 2: Autumn 1984 and Spring 1985
Individuals £10.50 (UK) £12.00 (overseas inc. Eire)
$14.50 (USA) $17.75 (Canada)
Institutions £24.50 (UK) £29.50 (overseas inc. Eire)
$40.00 (USA) $49.00 (Canada)

Orders with payment, or sample copy requests to: **Iris Taylor**
Basil Blackwell, 108 Cowley Road, Oxford, OX4 1JF England

Basil Blackwell for the
Social Philosophy & Policy Center

The Philosophical Quarterly

Vol. 35 No. 140 July 1985

Edited in the University of St Andrews by an Editorial Board
Chairman of the Board of Editors: LESLIE STEVENSON
Executive Editor: JOHN C. HALL *Reviews Editor*: PETER CLARK

Against Holism — *A. Weir*
Supervenience, Realism, Necessity — *I. McFetridge*
Three Accounts of Promising — *R. Downie*
Laws of Nature and Cosmic Coincidences — *J.J.C. Smart*

DISCUSSIONS
On Vico — *I. Berlin*
Berlin on Vico — *P. Zagorin*
Physicalism and Phenomenal Qualities — *E. Conee*
Names and Possible Objects — *M. Cook*
Hacking on Frege — *M. Dummett*

Subscriptions for 1985

Annual Subscription	U.K.	Overseas	USA	Canada
Institutions	£26.00	£32.50	£57.50	£71.50
Individuals	£10.25	£12.50	£26.00	£30.00

Orders, remittances and enquiries should be addressed to: Journals Department,
Basil Blackwell Publisher Ltd., 108 Cowley Road, Oxford OX4 1JF, England.

The Journal of Philosophy

Subscriptions at $20.00/year; $15.00 to students
$25.00/year to libraries

Students' price extended to retired and unemployed philosophers. Published bimonthly 1904 to 1976, monthly thereafter. Complete volumes and all separate issues available back to January 7, 1904 (volume I, number 1); prices as follows:

BACK VOLUMES (from the first to the penultimate year)
32-page issues, $1.50; 64-page, $3.00; special issues, $4.00.
Volumes (unbound, volume indexes included), $30.00 each.

CURRENT VOLUMES (this year and last)
64-page issues, $2.00; special issues, 96-page or over, $3.00.
Volumes (unbound, volume indexes included), $25.00 each.

ALSO AVAILABLE
Cumulative Fifty-year Index, 1904–1953; articles classified by subject and author; 452 p.; cloth, $12.00. Ten-year Supplement, 1954–1963: 98 p.; cloth $3.00, paper $2.00.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027